LANDSCAPE ANTHROPOMETRICS: A MULTI-SCALE APPROACH TO INTEGRATING HEALTH INTO THE REGIONAL LANDSCAPE

A Dissertation
Presented to
The Academic Faculty

by

Arthi Vijayanagara Rao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of City and Regional Planning

Georgia Institute of Technology
AUGUST 2016

LANDSCAPE ANTHROPOMETRICS: A MULTI-SCALE APPROACH TO INTEGRATING HEALTH INTO THE REGIONAL LANDSCAPE

Approved by:

Dr.Catherine Ross, Advisor
School of City and Regional Planning
Georgia Institute of Technology

Dr. Brian Stone
School of City and Regional Planning
Georgia Institute of Technology

Dr. Nisha Botchwey
School of City and Regional Planning
Georgia Institute of Technology

Dr. Adjo Amekudzi-Kennedy
School of School of Civil and
Environmental Engineering
Georgia Institute of Technology

Dr. Gonzalo Prokopec
Department of Environmental Sciences
Emory University

Date Approved: July 06, 2016

Dedicated to Sameer Vittal

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Catherine Ross for being my guide and mentor throughout my doctoral endeavors. She has always inspired me not to shy away from breaking new ground, no matter how daunting the task. I also wish to thank my dissertation committee for their invaluable feedback and support throughout the dissertation writing process. Their insights made this undertaking so much stronger, both from a conceptual and methodological standpoint.

Thank you to my family for walking with me through this journey. I owe a debt of gratitude to my husband, Sameer, for being there when I was absent, for keeping me focused, grounded and simply for listening. My son, Abhay, is a constant reminder of why I strive to do better.

Last but not least, I wish to thank my colleagues at the American Cancer Society for cheering me on to the finish line.

Most of all, thank you to each one of you for believing in me. I will remain forever grateful for your faith, guidance and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The explication of “healthy places” is currently a discipline-specific endeavor, fragmented along three primary axes. The first dichotomy exists in the anthropocentric vs. biocentric philosophies to defining healthy places. The second rift is evident in the reductionist methods and metrics employed to evaluate contextual impacts on human health. The third gap is with respect to scale. While there is abundant research investigating health and the built environment at the neighborhood scale, connections at the regional scale remain largely unexplored.

This research creates a consistent, scalable approach for incorporating health considerations into regional land planning for metropolitan areas. A prototypical framework is presented for the Atlanta metropolitan region. Determinants of healthy places from Social/Landscape Epidemiology, Urban Planning and Landscape Ecology are incorporated into defining the landscape and its associated pattern metrics. Key research objectives are to — 1) provide a new method to measuring urban form and health relationships through the use of landscape metrics 2) analyze urban form to understand configuration, mix, spatial distribution and proportions of land uses and socioeconomic factors that can support better health outcomes. Methodologically, this research examines associations between landscape patterns at nested scales (county and tract) with health outcomes measured by mortality rates across chronic conditions such as cancer, diabetes and heart disease. Two primary research questions are explored— 1) Are landscape patterns significant determinants of mortality rates? 2) At what scale do landscape patterns matter for reduced mortality rates?

Landscape Pattern metrics are generated using GIS software. Random Forest, Hierarchical Clustering and other classification techniques are used to identify preliminary landscape signatures and associations. Hierarchical impacts of regional determinants and tract-level landscape patterns on local health outcomes are examined through multilevel logistic modeling. The aim is to present a succinct set of landscape metrics to inform land use planning for healthy communities. The conceptual framework developed and investigated in this dissertation also presents a consistent, scalable method that can be used for multiple applications including Transportation Planning and sustainable Comprehensive Planning at multiple scales.

CHAPTER 1

INTRODUCTION

Problem Statement

The topic of built environment and health currently occupies a formal position of academic inquiry within the fields of public health and planning alike. The reconvergence between planning and public health is cemented and growing. This research paradigm is firmly embedded in the socioecological model or the idea that the health status of an individual is simultaneously produced by individual biology and the surrounding physical, social, cultural and political environment. Collectively termed the "social determinants of health," these are known to influence individual health as well as the differential distribution of health outcomes among populations (Jackson et al, 2013; Corburn, 2012; Berrigan and McKinno, 2008; Corburn, 2007a; Corburn, 2007b; Corburn, 2005; Northridge et al, 2003).

While there is abundant research investigating the connections between health and the built environment at the neighborhood scale, the connections at the regional scale remain largely unexplored. Thus, the main objective of this research is to create a multilevel framework for incorporating health considerations into regional land planning for metropolitan areas. This framework is created by expanding the socioecological model of health in three ways by— 1) incorporating relationships/determinants of healthy places as defined by disciplines such as Social/Landscape Epidemiology, Urban Planning and Landscape Ecology; 2) creating a new, more holistic methodology for measuring built form and validating its utility for research on "healthy places"; 3) explaining hierarchical impacts of regional land use patterns on local health outcomes through multilevel modeling. This approach and framework is explored for the Atlanta region as defined by the Atlanta Regional Commission (ARC) as the jurisdiction of the Metropolitan Planning Organization).

The motivation behind this research mainly stemmed from the Health Impact Assessment of PLAN2040— the regional land use and transportation plan created for the Atlanta region by the ARC. A recurrent theme during project meetings was the need for a core set of metrics that could connect land use and health and tracked over time. The literature revealed a fairly narrow approach to investigating links between land use and health (focused on walkability and obesity). On the other hand, regional approaches to investigating environmental correlates of infectious, vector-borne diseases are not extended to chronic diseases.

Serving as a philosophical, conceptual and methodological bridge, this dissertation brings together multiple current approaches from the fields of Urban Planning, Epidemiology and Urban Ecology. The framework developed here can be characterized as the beginnings of a regional "surveillance system", one that can be spatially and temporally extended to chronic and infectious diseases alike. It's closest analogous twin is the concept of "signatures" used routinely in the remote sensing world to detect landscape patterns. The idea is to create metrics that measure urban land use signatures as representations of healthy and unhealthy places. From an Epidemiological perspective, it investigates a much wider set of land uses and their role as exposures in the complex pathways to disease causation.

Several disciplines contribute to the definition of healthy places, however, the man-nature duality is apparent. Social Epidemiology and Urban Planning emphasize quantitative and qualitative aspects of sociocultural systems that impact the ability of the built environment to support human health (Corburn, 2007b; Williams et al, 2003; Williams and Collins, 2001). Landscape Epidemiology and Urban Planning (to a lesser extent) incorporate natural systems but the focus is on nature as a habitat for disease vectors or nature as a means to supporting healthier human societies (Lambin, 2010; Steiner et al, 2013, Steiner, 2011). On the other hand, ecologists characterize humans as "anthropogenic disturbance regimes" that negatively impact the health of natural ecosystems.

Sub-disciplines such as Landscape and Urban Ecology attempt a more integrated approach to human and natural systems. They provide theoretical and methodological tools to quantify the impact of changing landscape patterns on ecological systems. Landscape patterns are operationalized through the use of "landscape metrics". Land cover data is analyzed to understand the distribution and spatial arrangements of patch types such as forest, wetlands, impervious surfaces, agriculture, and other land and water types. Metrics typically quantify attributes such as shape, fragmentation/clustering, diversity, density of the different patch types within specified boundaries. Landscape ecological approaches have recently been discussed to develop and promote sustainable land use (Alberti, 2007; Leitao, 2006). While most prior studies tend to be descriptive or document the impact of human activities on landscape patterns and processes, connections to human well-being are largely absent or indirect (Alberti, 2003)

While Landscape ecological studies use land cover data as an approximation of land use, this dissertation uses a hybrid of land use and land cover data more representative of human activity patterns. The rationale for including land use data stems from both theoretical and methodological necessities. Land cover data represents material characteristics of the Earth's surface such trees, soils, asphalt, water, grass, etc. but does not directly reveal information about the cultural and socioeconomic uses of land. Theoretically, land cover data is more suited to research on zoonotic infectious disease transmission, vectors and their environmental (natural) habitats.

Land use data on the other hand is more representative of socioeconomic characteristics and other human uses of the land that might not coincide with land cover data. It is a physical manifestation of economic, cultural, political and other aspects of human activity (Vargo et al, 2013; Comber et al, 2008; Jensen, 2005; Brown et al, 2000). This dissertation focuses on chronic disease outcomes which are closely linked to human activity as determined by socioeconomics and land use patterns.

The framework developed in this dissertation also extends the built environment and health research to regional scales. Obtaining precise parcel-level land use data for the 21 county Metro Atlanta region poses data challenges. The classification of land uses also varies from county to county. In order to standardize the documentation of land uses for regional planning purposes, the Atlanta Regional Commission (ARC) publishes a hybrid land use/land cover dataset that interprets information from aerial photography as well as parcel level data wherever available (detailed description of dataset is available in chapter 4). This also demonstrates and benchmarks the utility of such regional datasets for incorporating health into decision-making. The data and landscape metrics are generated from the land use classifications in the LandPRO dataset.

From a methodological perspective, there is a widespread agreement that the socioecological model represents an intricate web of causation with simultaneously interacting agents. However, the execution of research takes on a more reductionist approach where outcomes and determinants are treated as discrete systems, often ignoring comorbidities and cumulative causation.

Landscape metrics are generally classified into two broad categories—composition and configuration. With respect to measurement, the distinction between compositional and configurational metrics is an important concept for this dissertation. ***Compositional metrics*** are calculated without reference to spatial attributes, such as placement or location of land use patches within a mosaic. Examples of compositional landscape metrics include those that measure variety and abundance of land use types in a landscape (a user-defined boundary such as a tract or a county). Here, only one value per landscape is created (for example, Shannon diversity index value for a county).

Configurational metrics refer to the spatial character and arrangement, position, or orientation of patches within the class or landscape. Examples include metrics that measure patch size distribution and density as well as proximity/interspersion of similar land use patches. Here, an independent set of metrics is created for each land use type within the landscape (McGarigal and Marks, 1995).

Traditional approaches to measuring urban form for planning (and subsequently for health) research utilize compositional approaches that inform on the presence and absence of features. For example, sprawl metrics indicate the presence or absence of land use diversity and density. However, there is little information regarding the quality of diversity and the spatial arrangement of constituent land uses that comprise this diversity (configuration). While beneficial and appropriate for transportation research, their utility for health research has not been fully justified. Even for transportation research, researchers has shown that travel behavior vary widely depending on the configuration of land uses, suggesting that composition metrics are insufficient (Cervero and Murakami, 2010; Hess et al, 2006).

Frenkel and Ashkenazi (2008) state that density measures tend to be the most commonly used to characterize sprawl. From a theoretical perspective, density is not a critical consideration for all health outcomes. More specifically, population density is important for transmissions of infectious disease but its utility for chronic disease is mixed. While lower levels of sprawl (higher densities) might reflect characteristics of the built environment that encourage physical activity, they might not be appropriate for studies that directly measure health outcomes or for diseases where land use may operate on a path other than physical activity.

Berrigan et al (2014) show that in the case of cancer, mortality rates were higher in less sprawling areas, even for obesity-related cancer. This is one demonstration of why sprawl measures alone might be inadequate in characterizing built environment and health relationships. While density is not the sole construct in characterizing land use patterns for this dissertation, it is accounted for through mortality rates (dependent variables) and the different classes of residential land uses that are categorized by density.

We also know that the health outcomes of individuals in a poor community surrounded by rich communities can be different from those of individuals in a large cluster of deprived communities. Research has also shown that individual health is impacted by contextual levels of social fragmentation and deprivation (Stafford et al, 2008; Stafford and Marmot, 2003). This dissertation marks an important departure from existing approaches to understanding land use-health relationships— it incorporates the above mentioned notion of "pattern", analyzing the built environment as a composition and configuration of spatial relationships and their implications for human health outcomes. It highlights the importance of spatial arrangement in linking built environment and health implications**. The "landscape" in this dissertation is conceptually defined as a combined fabric of land use and socioeconomic patterns. Pattern analysis of this combined surface moves built environment and health research towards a more holistic methodological direction (Figure1).**

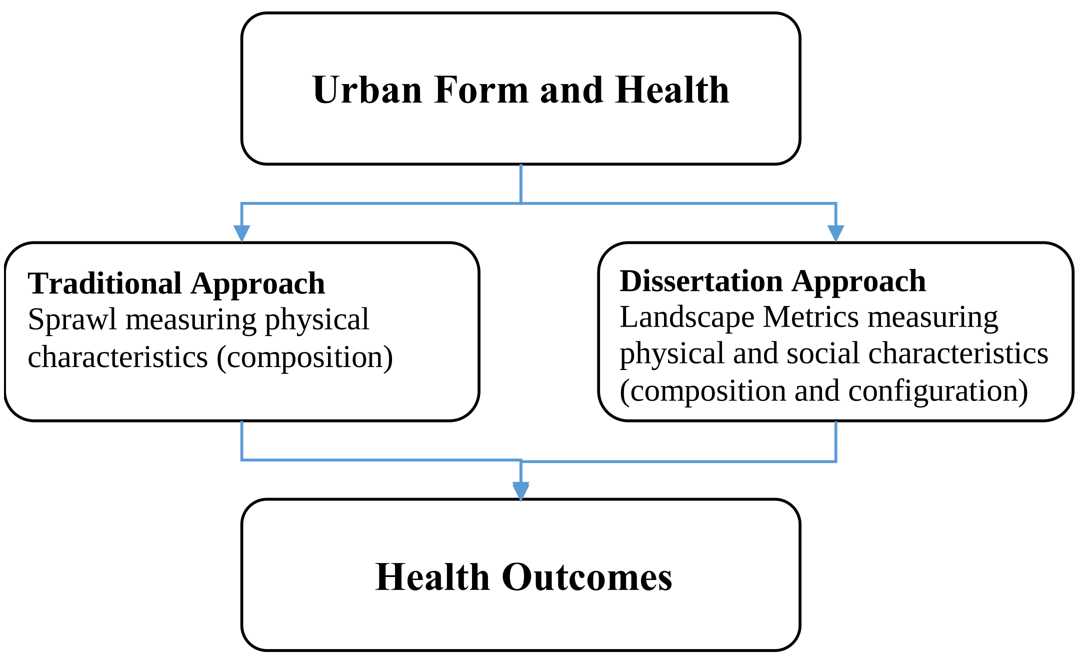


Figure 1. Defining the Landscape

This dissertation moves the needle from compositional approaches to built environment and health studies to a more holistic approach that includes composition and configuration. An important distinction is that it defines built environment as a landscape that incorporates physical and social characteristics (combined consideration of land use pattern and process).

The interconnectedness between natural and human ecosystems is highlighted by current phenomena such as climate change. Natural ecosystems provide ecosystem services critical to our well-being. Biodiversity and other such indicators of healthy ecosystems has been shown to be correlated with resilient places (human resiliency in the face of natural disasters) (Yang et al, 2013; Corvalan et al, 2005). Land use/land cover changes have also been implicated in the resurgence of infectious diseases. There is also increasing acknowledgement that ecosystems are hierarchical entities and regional processes influence local ones (Yarrow and Salthe, 2008). The use of land pattern analysis enables a more multidisciplinary, multiscale, holistic approach towards defining healthy places. It creates a common language for incorporating anthropocentric and biocentric priorities.

Landscapes present an entire spectrum of scales at which to recognize environmental patterns and their impact on human health. This coincides well with questions pertaining to urban planning and the “right” scale to plan for sustainable urban development (local, metro or megaregion?). Urban development patterns also correlate with health outcomes at different scales. For example, walkability is a neighborhood attribute related to obesity, a major risk factor for chronic conditions such as heart disease, diabetes, cancer, etc. (Feng et al, 2010; Josh et al, 2008; Lopez-Zetina, 2006; Ewing et al, 2003).

On the other hand, certain infectious diseases transmitted through zoonotic vectors such as lyme disease, dengue fever, etc. are influenced by regional characteristics of land use by increasing exposure to vector habitats (Gottdenker et al, 2014; Patz et al, 2004). There is a relative abundance of research that connects landscape metrics with the underlying processes of natural ecosystems (Ueemaa et al, 2009; Ueemaa et al, 2013). This dissertation uses that same language to connect with human health outcomes and processes. Theoretically, it forms a two-way shift in "ecosystem" and "socioecological" research where the two become more mutually inclusive. This can help with developing more balanced and sustainable policy and decision-making approaches for land use purposes. In the long term, this methodology can be used to explore the question "How can we use landscape metrics to create a more inclusive approach to the health of human and natural systems?"

Study Summary and Objectives

This research integrates approaches from Urban Planning, Social/Landscape Epidemiology and Landscape Ecology to defining healthy places. While these approaches are well developed within their disciplinary envelope, the paucity of interdisciplinary connections establishes a need for the research undertaken in this dissertation. This research serves as a pilot study to explore preliminary hypotheses and benchmark initial associations in an effort to create a roadmap for further detailed investigations.

Due to several feasibility considerations, the dissertation initially creates a prototypical framework for the Atlanta metro area. In particular, the study utilizes a census tract level dataset on health outcomes (mortality rates) made available by the Georgia Department of Public Health through special permission. This fills an important gap in the health and built environment literature where most studies utilize surveys (such as BRFSS and NHANES) where individual health is geocoded only up to the county level.

LandPRO 2010, a land use dataset derived from land cover and ownership data, presents a unique opportunity to compare land use patterns consistently across several counties. It is challenging to obtain such data for other urban regions that are classified in a manner consistent with LandPRO. The framework developed for Atlanta can then be used as a prototype applicable to other regional scales and places as well. While mortality rates are used in this research, the framework is set up to be easily adapted to other health outcomes data where available. Landscape Ecology offers a particularly appropriate, replicable theoretical and methodological approach for this exploration. In addition to health outcomes, the land use framework developed here can easily be extended to other facets of Urban Planning research such as Transportation and Economic Development Planning.

An analysis of the mortality rates reveals the primacy of heart disease, lung cancer, chronic obstructive pulmonary disease (COPD) and diabetes as predominant causes of death in the Atlanta area. Accordingly, the dissertation investigates relationships between landscape patterns and these health outcomes.

In summary, the dissertation achieves the following objectives:

- Provides a relatively elegant approach to guiding healthy land use development at the regional scale that optimizes health outcomes and determinants.
- Introduces a holistic approach to land use planning by—1) Understanding the contributions of landscape pattern (cumulatively defined by urban form and socioeconomic factors) to the impacts on human health; and 2) Exploring the utility of landscape metrics (as used in landscape ecology) in enabling healthier land use planning (both through visualization and measurement). Ideally, the research will identify a key set of landscape metrics that can be easily applied at different scales in land use planning 3) Understand the scale at which these patterns matter (Metropolitan region, county or census tract)

Three primary research question are explored to fulfill the above stated objectives—

1) Are landscape patterns important determinants of human health?

Urban Planning examines associations between urban form and health outcomes/determinants, albeit in a limited way. For example, macroscale studies examine correlations between urban sprawl, obesity and associated health outcomes at the metropolitan and county levels (McCann and Ewing, 2003; Ewing and Hamidi, 2014). At the neighborhood scale, various land use and urban design characteristics that support health determinants such as physical activity, access and social capital are studied through the compilation of metrics such as walkability indices and other proximity measures. However, several of these indices are compiled with little knowledge of objective thresholds, interactions between land uses and their proportions. The evidence linking built environment and health outcomes is mixed. Some studies show weak or insignificant links when controlling for socioeconomic characteristics. This dissertation strives to establish objective thresholds and metrics at macro and micro scales through the use of a consistent methodology. In order to explore this question in greater depth, the following related questions are investigated:

i. *Are landscape patterns significantly correlated with health outcomes?*

ii. *Do social patterns have a stronger association with health outcomes compared to landscape patterns or is there an interactive effect?*

Exploratory analysis is used to uncover relationships between landscape pattern variables associated with health outcomes at census tract and county scales. This includes the use of correlation analyses and hierarchical clustering studies. Confirmatory models are then used to further extract statistically significant predictors.

2) ***How is land use mix and spatial distribution (composition and configuration)of landscape components (land use and socioeconomics) associated with health outcomes?***

Holistic approaches to understanding the urban landscape as a mosaic composition and combination of interacting parts is largely unexplored. For example, we know that a mix of land uses and higher densities promote walkability. However, what are those thresholds and how do they impact health positively or negatively? For example, access to greenspace is often calculated with simplistic metrics such as number of park acres per unit of population. However, there is little attention paid to the quality of the greenspace itself. Is it better to have one large park or smaller parks distributed through the neighborhood and county? What are the optimal distances and relative juxtapositions between parks and other land uses? Which configurations are better for physical activity vs. which ones are better for ecosystem functions? What are the optimal set of landscape metrics to use that parsimoniously explain these relationships? While measuring these microscale phenomena is beyond the scope of this dissertation, the study investigates preliminary associations between land use types (parks included) and manifestations of physical inactivity (heart disease and diabetes mortality).

3) At what scale do landscape patterns significantly impact human health outcomes?

Incorporating health metrics into regional land use planning is gaining national significance in the United States. This is highlighted by the efforts of Federal organizations (FHWA), Metropolitan Planning Organizations and other regional planning entities are consciously using health as an important measure of plan quality as well as important criteria to guide project selection and investment.

This research investigates the scale at which landscape patterns matter. For example, it is hypothesized that heart disease and diabetes mortality might be more influenced by local patterns (census tract level-since lifestyle factors such as physical inactivity and poor food access are significant risk factors). On the other hand, lung cancer and COPD mortality might be more dependent on patterns at the county or larger regional scale (linked to environmental pollution). Compared to existing approaches, landscape metrics provide a consistent methodological language across scales, enabling comparisons of different urban form typologies. Most importantly, at what scale do we start seeing a correlation between these ordered patterns of urban elements and health outcomes/determinants?

Chapter Outline

The dissertation begins with a review of the relevant literature that informs this research as well as gaps in the literature that help establish the research questions, objectives and significance of the work. Chapter 1 introduces the background and relevance of the study. Chapter 2 includes a review of theoretical and methodological constructs from the contributing disciplines identified above, their linkages and the contributions that the dissertation makes in adding or strengthening those linkages.

Chapter 3 lays out the research framework that serves as a roadmap for this investigation. It includes the description and rationale for the geographical location of the study and formally states the research questions and hypotheses. The chapter then outlines the analytical framework. This includes operationalizing the constructs stated in the research questions (dependent and independent variables) as well as a description and sequencing of the quantitative analysis that are carried out to answer the questions (analytical framework).

The next chapter (chapter 4) dives into details of the raw datasets, the various aggregations and transformations applied in creating the variables and their theoretical relevance to the study. The first section describes the raw mortality data received from GaDPH, selection of the top causes of mortality and computation of the adjusted mortality rates. It concludes with a study of the spatial distribution of the mortality rates and their final conversion to the binary dependent variables. The next section provides details on the publicly available LandPRO dataset published by the Atlanta Regional commission, the land use categories and their definitions. The Patch Analyst software, the various landscape metrics and their definitions are introduced. Finally, the rationale for selecting a subset of land uses and their associated metrics are laid out, culminating in the Level 1 and Level 2 land use indices used as the primary independent variables. Section three describes the rationale for constructing the socioeconomic index, the Census datasets and variables used and methods used to aggregate this data. In addition to measuring neighborhood socioeconomic status, it represents neighborhood deprivation and other contextual/neighborhood attributes not directly captured in the land use dataset.

Chapters 5 discusses the various types of confirmatory data analysis executed in this dissertation, for Lung Cancer and CODP as a demonstration of the framework and methods. It includes the process of variable selection, logistic and multilevel modeling as well as an interpretation of results for each of the disease types. Multilevel modeling is utilized to ascertain the association between landscape metrics and health outcomes at two levels—1) census tract 2) county. Chapter 6 brings the entire document together to discuss the implications of the work for the study region as well as its potential applicability to other regions and purposes.

Methodologically, this dissertation utilizes the concept of landscape metrics to investigate connections between urban form and human health outcomes/determinants. Philosophically, it adapts an urban ecology approach where humans are an integral part of the ecosystem (anthropogenic biome) and human health outcomes are of primary interest. From a policy perspective, the aim is to describe typologies of healthy urban form (landscape pattern) that incorporates biocentric and anthropocentric considerations.

The framework developed in this dissertation also extends the built environment and health research to regional scales. This poses data challenges to obtain precise parcel-level land use data for the 21 county Metro Atlanta region. The classification of land uses also varies from county to county. In order to standardize the documentation of land uses for regional planning purposes, the Atlanta Regional Commission (ARC) publishes a hybrid land use/land cover dataset that combines information from aerial photography as well as parcel level data wherever available (detailed description of dataset is available on pg.40 in the methodology section). This also demonstrates and benchmarks the utility of such regional datasets for incorporating health into decision-making. In this proposal, the data and metrics will be called “land use” going forward.

Methodologically, landscape metrics provide a practical and holistic approach to measuring the composition and configuration of landscape elements. They help quantify aspects of spatial distribution and juxtapositions which help us understand contextual relationships. They also form a very practical toolkit as metrics are easily generated through freely downloadable software packages available as add-ons to existing GIS platforms (Patch Analyst, FRAGSTATS).

Individual and aggregated socioeconomic characteristics have nullified the effect of the built environment on health in some studies. In Landscape Ecology, landscape metrics strictly evaluate land cover/land use relationships. However, social patterns impact health outcomes as well. For example, studies in social epidemiology characterize the spatial heterogeneity in residential segregation as an important indicator of health outcomes. Landscape elements or patches are analyzed in this study in combination with their socioeconomic characteristics as well. This tests for the primacy of physical and social patterns simultaneously.

Land use planning forms the foundation in determining human activity patterns as well as land conservation— the critical elements of sustainable development. From a policy perspective, this dissertation provides a comparative glimpse of how landscape patterns relate to human health as compared to established concepts of ecosystem health using the same metrics. Policy guidelines will be developed to inform land use planning at regional and local scales. Ultimately, the aim is to create a policy and visualization toolkit that regional planning agencies can use in a balanced approach towards creating healthy places.

CHAPTER II

LITERATURE REVIEW

The socioecological model of Public health

The socioecological model of public health provides the most comprehensive framework within which to understand the relationship between contextual factors and individual health outcomes. Health of the individual is explained as a constant interplay between individual behavioral/genetic/socioeconomic characteristics as well as hierarchical levels of environmental factors such as the cultural, social, built and political environments. While this model provides a generic framework that delivers the message that health is not solely determined by individual factors, the contextual factors vary based on the discipline that is adopting it. The hierarchical levels of contextual influences as enunciated by various disciplines are also ripe with explanatory potential.

Urban Planning, Land Use and Health

Planning healthy communities is a central tenet of sustainable urban planning. The connections between built environment characteristics at the neighborhood scale and their impact on health determinants (walkability, access to services, and access to healthy foods) as well as health outcomes (disease morbidity and mortality) are extensively explored in the planning and public health literature. However, the connection between health and regional land use planning is a topic of very recent inquiry. Irrespective of scale, land use planning forms the foundation of planning healthy places (Lejano et al, 2012; Leitao et al. 2006; Jackson et al, 2002). Land use patterns are a result of land use planning practices and directly impact the distribution of human activity patterns as well as natural ecosystem dynamics.

The literature on land use planning and health can be broadly classified into two categories — 1) studies that research the associations between land use and health outcomes as mediated through health determinants 2) studies that focus on the direct associations between land use and health determinants. In the first category, aspects of land use composition are studied in relationship to chronic disease risk factors and outcomes such as obesity, heart disease, diabetes. Research in the second category tends to focus on the ability of land use attributes to influence healthy behaviors (physical activity) or serve as sources of environmental exposures (air and water pollution).

In the first category, the connection between sprawl and health outcomes forms the most prominent research framework (Ewing and Hamidi, 2010; Frumkin et al, 2004; Ewing and McCann,). In the urban planning literature, the concept of sprawl is a well-established technique of measuring land use/development patterns. While several definitions of sprawl exist, researchers agree that sprawl is a multi-dimensional measure which is fairly complex (Ewing, Pendall and Chen, 2002; Galster, 2001; Banai and DePriest, 2010). For the purpose of investigating health outcomes, the framework created by Ewing and Hamidi (2014) is the most thoroughly explored. Similar to most studies, sprawl is characterized as a combination of 4 factors— development density, land use mix, activity centering and street accessibility. While all the factors measure aspects of urban development patterns, each of the factors hold implications for their impact on health outcomes. Compact areas were found to impact the social determinants of health by providing greater economic mobility and opportunity, lowering household expenditure on housing and transportation costs and providing greater transportation choice (accessibility). Compact areas showed a direct association between longer life expectancy, lower BMI and reduced rates of high blood pressure and diabetes. The study also provides additional insights into the relationship between land use characteristics and health. For example, the authors hypothesize that the degree of sprawl impacts life expectancy through four mediating factors, namely, obesity, traffic fatalities, air pollution and violent crime.

The 2014 study is an update on the previously developed sprawl index (Ewing, Pendall and Chen, 2002; McCann and Ewing, 2003). Ewing and Hamidi (2014) find that sprawl factors are found to have significant relationships with obesity and chronic conditions even after controlling for individual socioeconomic characteristics. These findings are consonant with the earlier sprawl study where the authors McCann and Ewing (2003) found that residents of sprawling counties have higher BMIs compared to residents of more compact counties. While the 2014 sprawl study provides several important updates such as the use of path analysis and multilevel modeling in establishing potential causal relationships, some weaknesses of the study are evident. From a theoretical standpoint, it establishes the connection between place, health determinants and health outcomes. The study utilizes publicly available health data geocoded up to county level. This prevents the study of neighborhood level associations that might be more evident at the tract level. The sprawl measures used in the study are purely compositional, i.e., they provide little insight into the varying spatial distributions and juxtapositions (configuration) of land uses that might vary between counties that have similar sprawl metrics. The factors also provide an incomplete characterization of land development patterns. For example, the land use mix factor (most relevant for this study) only examines jobs/housing balance. Sprawl metrics capture averages and provide limited insight into the actual physical planning of places through size and placement thresholds for land uses.

In addition to the framework described above, several other studies explore the relationship of urban sprawl and obesity at regional and neighborhood scales (Mackenbach, J. D., 2014; Zhao and Kestner, 2010). The results are inconsistent and vary based on the methods and measurements used as well as the geographic location of the study itself. For example, some studies use population density as a crude proxy for characterizing sprawl. While there is a significant body of literature that shows a positive association between sprawl and obesity, these results get negated when factors such as self-selection and other such explanatory variables are incorporated into models (Kostova, D., 2011; Eid et al, 2008; Plantinga and Bernell, 2007). Specific weaknesses pointed out in the studies include a wide variability in the definition of sprawl itself and the use of aggregated indices that offer limited explanatory potential between form and process. More specifically, the sprawl index does not examine the impact of different types of land uses. With respect to scale, Lathey et al (2009) offer a larger commentary on the characterization of urban form and health outcomes. They state that sprawl measures are also computed at the regional scale (metro, county, state), offering limited explanations of neighborhood conditions. In order to counteract these issues, they propose a method that correlates disaggregated measures of sprawl to disease clustering at the block group level. An interesting outcome of their study is that land use mix (particularly the quantity of green space) and socioeconomic characteristics were the strongest predictors of disease occurrence. Land use patterns (as measured by sprawl) have also been studied for their association with other chronic diseases such as cancer and heart disease with their relationship being mediated by health determinants such as physical activity, access to food, access to healthcare, etc. (Berrigan et al, 2014; Griffin et al, 2013; Gregson, 2011; Sallis, J., n.d.).

While most studies explain pathways between urban form and health through their ability to support health determinants, some studies have shown a direct association between urban form and health outcomes even after controlling for socioeconomic conditions (Lathey et al, 2009; Stockard et al, 2004). However, the overall conclusions about the primacy of built environment or socioeconomic status in determining health status is a prominent debate in the literature. Some studies show differential gradients in the impact of the built environment based on socioeconomic and demographic conditions (Bodea et al, 2009; Saxton-Ross, 2009).

Research in public health and urban planning documents the impact of land use planning on several health determinants. Most significantly, walkable neighborhoods characterized by diverse land uses (land use mix), high street connectivity, walkable/pedestrian-oriented destinations are often associated with higher levels of physical activity and reduced air pollution (due to less automobile use, lower VMT) even after controlling for socioeconomic variables (Frank et al, 2006; Frank and Engelke, 2005; Frank et al, 2004; Frank and Engelke, 2001) This research has direct implications for obesity related chronic conditions such as heart disease, cancer as well as those related to environmental pollution such as asthma and other respiratory illnesses. The intense focus on walkability and neighborhoods is largely motivated by the criticality of the obesity epidemic and ecological approaches towards its prevention.

Research results on the impact of land use mix (LUM) on walkability and health itself are mixed. Some studies show significant relationships between LUM, walkability and physical activity (Duncan et al, 2010; Frank et al, 2004; Christian et al, 2011). However, these tend to be more simplistic models that use linear or binary logit regression. More sophisticated models that incorporate multinomial logit and multilevel methods present a mixed bag of results. The effect size of LUM on walking behavior is greatly reduced after controlling for socioeconomic variables and has a differential effect based on baseline BMI status. However, it is interesting to note that other built environment attributes such as residential density and even social patterns of residential segregation are revealed to be significant variables (Bodea et al, 2009; Saxton-Ross, 2009). Handy et al (2008) move closer to establishing causal relationships by controlling for self-selection bias in their study design.

Christian et al (2011) offer specific critique with respect to the concept of LUM and walkability. They state that "At present there is no conclusive evidence on what aspects of land use are most important to encourage different types of walking and physical activity". Entropy based formula (Shannon's diversity indices) are most commonly use to operationalize LUM which are then incorporated into walkability indices. There is wide variation in which land uses get included in LUM calculations as well as in their aggregation. Researchers often use subjective judgment in their selection. Shannon's diversity index is often used interchangeably with Shannon's evenness index, leading to erroneous conclusions.

The influence of urban form on travel behavior is extensively documented in the transportation literature. A prominent framework is the one developed by Ewing and Cervero (2010) which identify the 5 “ds” of the built environment associated with travel choices. These are density, diversity, design, destination accessibility and distance to transit. *Density* represents population, housing or employment density and the variable chosen depends on the purpose of the study. *Diversity* is an indication of the variety of land uses present and is represented through entropy measures. *Design* includes characteristics of the street network that support active transportation. *Destination* accessibility measures proximity and ease of access to trip attractors. *Distance to transit* measures access to transit stops from residences or workspaces.

Hess et al (2001) provide a critique of traditional approaches used to measure land use characteristics for transportation research. This critique is corroborated by Manaugh and Kreider (2013) where they state the shortcomings of purely compositional measures of land use mix such as entropy index. Entropy measures do not differentiate between types of land uses, do not consider interactions or spatial arrangements of land uses and are unaffected when the proportions of land uses are reversed. As a consequence, two very different spatial configurations of land uses can possess the same entropy score. The authors propose a more nuanced methodology that introduces metrics from Landscape Ecology which captures compositional and configurational characteristics. This is an important piece of research that bolsters the claims made in this dissertation for studying a similar, detailed and easily scalable methodology for built environment and health research.

In reflecting on the methodological challenges associated with establishing connections between built environment and health, Frank et al (2006) document reductionist approaches as a major weakness. Studies tend to focus on a single health outcome such as obesity or a single determinant such as air quality. Multilevel studies offer greater external validity in understanding contextual determinants of health when they examine individuals nested within counties (this is due to the nature of publicly available data). However, the scope of neighborhood level studies is limited as they examine a very small sample of individuals nested within a few neighborhoods (extensive need for data collection).

From Land Use to Landscape: Moving towards a more holistic conception of urban form

The literature review above, explains how built environment and health studies are narrowly focused on obesity and walkability relationships at the neighborhood scale. Natural ecosystems and their interactions with human health are also conspicuously absent in the conversation. However, the socioecological model of health posits a far more complex understanding of health at multiple scales. Towards this purpose, the next section reviews other disciplinary frameworks hereby largely unexplored in urban planning research. These frameworks make important contributions to the understanding of health and place, at different scales and also address the combined concept of "ecological health" —one that is inclusive of natural and human systems (landscapes). They also extend the concept of land use to landscape in that they integrate vertical (topological) relationships with chorolological (horizontal) and contextual relationships.

The objective is to productively use elements of these frameworks in creating a complex, comprehensive and multi-scale approach towards informing the creation of healthy places.

Landscapes and Health

Defining landscapes

In a philosophical approach to understanding landscapes, Von Maltzahn (1994, p.109) presents two distinct definitions of the meaning of the word "landscape". The objective definition—"landscape as the actual land as it is in its own composition, regardless of whether we as experiencing persons are present or not" alludes to an analytical framework of dissecting the landscape into its constituent parts. The subjective definition—"landscape as the experiential space of everyday life which requires the presence not only of the actual land but also of ourself with our particular point of view" suggests a more perceptual process of the construction of meaning through individualized engagement with the surrounding environment. Geographers such as Jay Appleton have further theorized the relationship between the objective and the subjective. For example, how does the objective structure of the landscape influence our perception of it? Habitat theory elucidates qualities such as "prospect" and "refuge" that influence the development of sense of place and aesthetic appreciation of the landscape.

While both definitions are equally important from the viewpoint of "health and place", this research will emphasize an investigation of the former. However, it is integral to provide a framework for ongoing research that connects the two. A future follow-up research agenda might ask the question "How do regional landscape patterns impact local landscape patterns and perceptions in the production of health outcomes?" Towards that purpose, a brief overview of the different disciplinary constructions, metrics and methods of health determinants in the landscape are explained in the following section.

The different dimensions of Healthy Landscapes: Disciplinary conceptions of landscape and health

This dissertation draws on the following primary disciplinary bodies of work in defining and understanding the connection between place, people and health—urban planning, landscape/spatial epidemiology, social epidemiology and landscape ecology. Common to all these disciplinary frameworks is that the production of health occurs through interactions between people, places and social/cultural forces. Each discipline then attempts to weave an explanatory "landscape" or spatial fabric of health determinants or health networks rooted in different disciplinary understandings of place.

The physical aspects of "place", also common to these disciplines, can be further expanded into a multilevel framework. Also prominent is the evolution from biomedical investigative approaches to a more holistic continuum that embraces well-being rather than the mere absence of disease. This approach is consonant with the adoption of the socioecological model, forming an important link between the built environment and public health in general. However, what differentiates each of these disciplines is their areas of focus as well as varying degrees of theoretical and methodological complexity. The next few sections of this chapter discuss these aspects, building a progressively comprehensive understanding of the relationships between land use, landscape and health.

Epidemiology

Epidemiology is “the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to control health problems” (McKenzie et al, 2012). An important aspect of Epidemiology is defining exposure— the factor or factors primarily associated with the health outcome of interest. In epidemiological studies the term “exposure” can be applied to the primary explanatory variable of interest along with confounders and effect modifiers associated with the health outcome being investigated. Three epidemiological sub-disciplines, namely, Landscape Epidemiology, Spatial Epidemiology and Social Epidemiology are further explicated in this literature review. They are specifically relevant because they align with the exposure variables of interest in this dissertation and offer sound conceptual, theoretical and methodological foundations for this research.

Globalization and urbanization contributing to climate change have led to significant changes in land use patterns. These changes have put humans in close contact with natural systems which also serve as vector habitats for zoonotic diseases. As a consequence, infectious diseases are making a comeback in geographies where the disease was hitherto non-endemic. This poses a new frontier of threats for public health, in addition to the epidemic increase in chronic disease.

“The theory behind landscape epidemiology is that by knowing the vegetation and geologic conditions necessary for the maintenance of specific pathogens in nature, one can use the landscape to identify the spatial and temporal distribution of disease risk” (NASA,n.d.). The landscape, defined as a complex ecosystem comprising natural and man-made elements (biotic and abiotic conditions), has attributes that potentially influence disease risk and incidence. The theoretical underpinnings of Landscape Epidemiology involve understanding these complex interactions between natural (for example, vegetation), geological (for example, soil) and cultural (for example, the built environment) attributes that support pathogens and zoonotic vectors in transmitting disease. Combinations and configurations of landscape characteristics can then be analyzed to identify the spatial and temporal distribution of disease risk. Examples of key environmental indicators used include elevation, temperature, rainfall, and humidity, vector/pathogen habitats and their development, activity, and longevity and their interactions with human systems (population densities, land uses, infrastructure and movement).

The landscape in this disciplinary field can be described as a surface comprised of different typologies representing various combinations of environmental characteristics. These typologies in turn represent differing potentials to harbor and transmit disease causing pathogens.

Spatial Epidemiology is a largely methodological field that finds applications in medical/health geography and landscape epidemiology. It is a combination of geospatial and statistical methods used to investigate distributions of disease, risk as well as establish correlations between disease incidence and suspected causative environmental factors. Analysis is done across time and space. Elliott and Wartenberg (2004) classify predominant areas of spatial epidemiologic inquiry as follows:

- Disease mapping
- Geographic correlation studies

- Disease clustering and their correlates

Through a combination of correlation and clustering studies, spatial interpolation techniques create landscapes or surfaces of risk. These models can be further developed for predictive purposes to study the relationships between disease rates and changes underlying determinants.

Social Epidemiology explores how the built environment is an expression of or reinforces social phenomena such as socioeconomic status, racism, social support and stress, thus mediating the relationship between Health and Place. Also examined in depth is the validity of constructs used to operationalize these qualitative variables and their subsequent consequences for research. “Ecosocial theory” or the “social-ecological systems perspective” analyzes population health on a continuum of biological, ecological and social factors (for e.g., cell, organ, organism/individual, family, community, population, society, and ecosystem). Krieger (1994) aptly characterizes the socioecological system of the production of disease and health as the “web of causation”. The focus is on understanding the complex and potentially non-reductionist interactions between all the levels and can be considered a more comprehensive theory that incorporates the first two. An important concept is “embodiment” or how the human body incorporates, biologically, it’s material and social environments and how these create different pathways between exposure, susceptibility and resistance. Examples relevant to urban planning include residential segregation /exclusionary zoning and neighborhood deprivation and how they affect health outcomes differentially among population sub-groups. Residential segregation leads to higher rates of economic deprivation among African Americans, who in turn can only afford to live in poorer communities without good access to healthy foods as well as higher exposures to environmental toxins such as lead paint in older homes and proximity to waste facilities. At the individual level, poor nutrition (high fat, high sodium foods and low vegetable consumption) in turn increases risk for obesity and both poor nutrition and environmental toxins are significant determinants of hypertension and chronic kidney disease.

Studies in Social Epidemiology have established independent effects of neighborhood level socioeconomic conditions above and beyond individual-level socioeconomic attributes (Messer and Kaufman, 2006; Messer et al, 2006) . Studies have also shown that socioeconomic conditions mediate the relationship between built environment characteristics and health outcomes. Area level measures of Socioeconomic Status (SES) are collectively operationalized as neighborhood deprivation indices. Based on the “collective resources” model, Stafford and Marmot (2003) explain the hypothesized link between health outcomes and neighborhood deprivation. Residents in non-deprived neighborhoods (higher neighborhood SES) have better health outcomes due to better access to collective resources (such as services, job opportunities, and social supports). Conversely, residents of deprived neighborhoods suffer the outcomes of disinvestment, resulting in “pathogenic residential conditions” (Williams, 2004). These are characterized by poor schools, poor access to healthy foods and healthcare, absence of parks/recreational spaces and high crime. Cumulatively, they influence health behavior and physiological responses, providing prime environments for the development of negative health outcomes. Neighborhood Deprivation has also shown to be correlated to multiple physical health outcomes such as cancer, diabetes, heart disease and asthma as well as mental well-being. In addition to a lack of resources, other mechanisms that mediate this association include chronic stress and lack of social cohesion/social support. (Sundquist, K., et al, 2015; Gale, S. et al, 2011; Richardson, E. et al, 2013; Stimpson, J., 2007; Gustafson, A., 2013; Barrington, W.E., 2014).

The process of redlining also creates disproportionate exposures to unhealthy land uses and other environmental toxins and poor housing quality for lower SES neighborhoods. Concentrations of joblessness, poverty and residential instability create chronic environmental stressors such as increased rates of violent crime. Research has shown that those that belong to low SES neighborhoods have poor access to medical care (Williams, 1999; Williams, 2001). Overall, exposure to mutually reinforcing factors such as low SES and racial discrimination produce persistent differentials in health outcomes between racial groups.

Here, the landscape is often referred to as a "web of causation" which tends to include an all-inclusive modeling of the different hierarchical levels of the environment. However, the actual causal mechanisms are considered elusive as all interactions between levels are hard to quantify. These models tend to be the most comprehensive, however, operationalization of individual levels are often simplified.

In reviewing the above frameworks for "health and place" research, the emphasis on anthropocentric considerations becomes conspicuous. However, in order to situate "healthy places" within the larger umbrella of sustainability, a more inclusive approach that encompasses both human and natural systems is necessary. As Coutts and Taylor (2011) point out, land conservation, green infrastructure and environmental protection are seldom considered in socioecological models as important components of upstream prevention measures. Further, Coutts (2010) proposes a “public health ecology” framework to illustrate the connections between landscape structure, land conservation and human health (Figure2). Advocating with the anthropocentric benefits of conservation is considered a powerful political tool to garner support for sustainable land use decision-making.

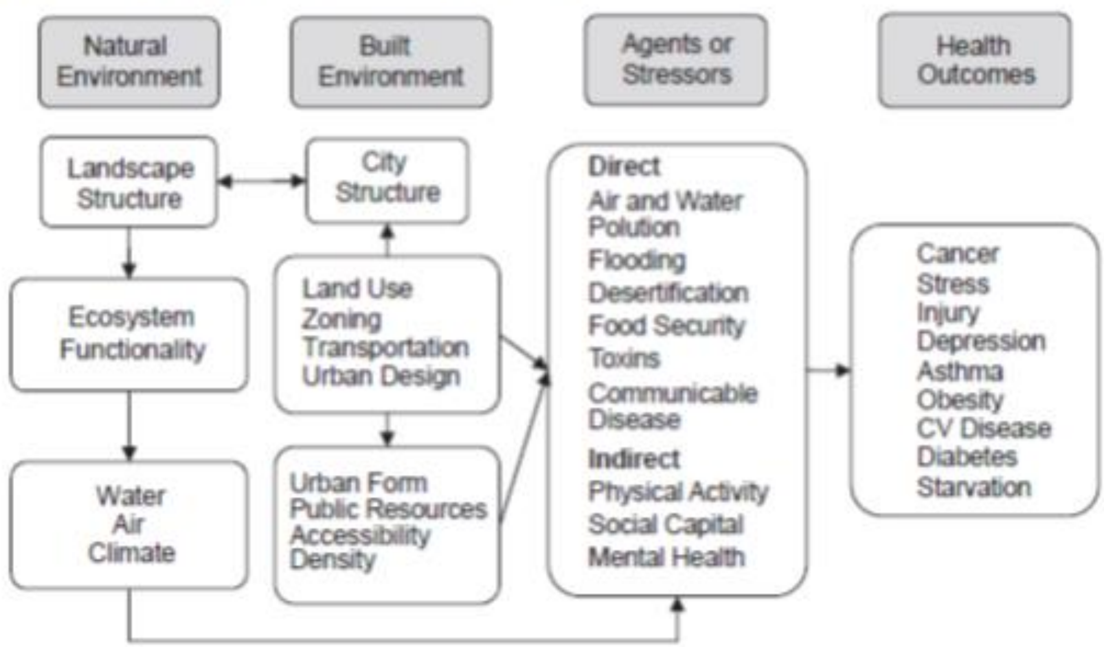


Figure 2. Framework of Public Health Ecology. Image Source: Coutts, 2010

Landscape Ecology: Pattern and Process

Normatively, sustainable urban development mandates a balancing act between natural ecosystems and human culture. As much as the interdependencies between natural and cultural systems are recognized, disciplinary practice emphasizes one or the other. These theoretical and methodological approaches to land use and natural resource management promotes a dichotomy between human and natural systems, inhibiting a truly integrated approach to sustainable development.

Health, be it health of ecosystems or human health, is often used to characterize sustainable systems (used as a performance measure of the system being studied). However, this duality between man and nature exists in the conceptualization and measurement of health as well (Jianguo et al, 2007; Hope et al, 2000). From a health perspective, the socioecological model of health emphasizes the relationship between individual health and the quality of an individual's environment. Earlier sections of the literature review reflected on different operationalization of environmental factors. While social and cultural factors are emphasized, the co-dependency between the natural environments and human health is slowly being recognized.

Landscape Ecology and Sustainable Urban Development

Alberti et al (2003) summarize the limitations of traditional ecological paradigms in assessing humans as integral parts of the Earth's ecosystems. Despite being a dominant species, humans remain excluded from ecological research and modeling. Emerging ecological paradigms such as those in Urban Ecology and Landscape Ecology do recognize humans as components of ecosystems. However, the focus of the questions needs to shift from "how do humans impact ecosystem functions and processes" to "how does the interaction between human and natural systems produce distinctive emergent characteristics of urbanized landscapes?" (Alberti, 2008).

While sustainability principles drive contemporary planning practice, ecological considerations are only recently being incorporated into planning. The primary emphasis is on the planning and management of green infrastructure as a vital component of preserving ecosystem functioning and services. From a larger perspective, the symbiosis between human and natural systems is highlighted by "the capacity of the earth to maintain and support life and to persist as a system". This urges us to think about larger relationships and interdependencies between land uses that extend beyond green infrastructure planning (Leitao et al, 2006).

More recently, the sub-discipline of Landscape Ecology provides a more three-dimensional approach to understanding ecosystem relationships. Zonneveld (1994) and Leitao et al (2006) describe three types of relationship between the attributes and elements of the Earth's surface:

- Topological (vertical)
- Chorological (horizontal)
- Geospherical (global)

Topological and chorological relationships are considered particularly important in their utility to inform sustainable planning practice. Planning (and plan-making) is an essentially spatial exercise. A place is a palimpsest of vertical and horizontal juxtapositions to form a complex system manifested in form and function. It is this understanding between point and area patterns that help bridge the connections between local and regional phenomena. While topological relationships measure the cumulative impact of vertical (local) elements, chorological relationships examine spatial heterogeneity among different landscape "patches" or "ecotopes" (Zonneveld, 1994; Leitao et al, 2006).

Landscape ecology offers us a conceptual articulation of "landscape" with regard to four different aspects—scale (geographical size), constitution (composition), measurement and temporal change.

The Landscape scale is considered a suitable scale for sustainable planning endeavors. By this, literature often refers to a regional scale, one with multiple nested hierarchical ecosystems interacting with one another. A standard size is not prescribed but definitely alludes to larger regions such as watersheds. However, from a species-centered perspective, there can be wide variation in size as each organism scales the environment differently based on habitats and behaviors (Leitao et al, 2006; McGarigal, 2014). Thus a range of scales is recommended for effective ecosystem management (McGarigal, 2014). This also considered a good approach to research a combined ecological approach of human and natural ecosystems and lends itself more easily to sustainability research. Another important consideration is that decision-making and planning for land conservation and natural resources often happens at the landscape or regional scale (Leitao et al, 2006; Wu, 2013).

With respect to composition, Landscape Ecology offers two important theoretical constructs. First, the concept of holism defined as "the total of abiotic and biotic phenomena and their interrelations in the three dimensional space on the earth's surface. It can be observed and recognized by its horizontal and vertical structure and its combination in the variation in attributes: atmosphere, rock, relief, soil, water, vegetation, animals and man." (Zonneveld, 1994, p.14). This approach forms an important conceptual bridge between anthropocentric and biocentric priorities in defining healthy places.

Second, the conception of landscape as a mosaic or collage of interacting ecosystems collectively characterized as landscape structure. The most common method of depicting landscape structure is the patch-corridor-matrix model. Patches are relatively homogenous nonlinear areas that are distinct from their surroundings. Corridors are linear land cover types which serve as connections and conduits for the movement of animals, plants, nutrients, etc. through the landscape. The matrix is the dominant land cover type that exerts control over the dynamics of the landscape. This is determined by area, degree of connectivity and continuity (Forman, 1995; Leitao et al, 2006).

The landscape elements (patches, corridors and matrices) are arranged in varying patterns across the Earth's surface. A primary concept in Landscape Ecology is that the composition and configuration of landscape elements (spatial patterns) determine and represent ecological processes. This relationship is analogous to the commonly used concept in the built environment disciplines, namely, the connections between form and function. Thus, in order to assess the "performance" or "health" of landscapes, it becomes necessary to quantify these spatial patterns. In Landscape Ecology research, this spatial heterogeneity is measured using landscape metrics. While the words **land use and land cover** are used interchangeably, it is predominantly land cover data that is used as the underlying data for this type of analysis. McGarigal (2014) identifies the following categories of metrics used in landscape research, commonly generated by software packages such as FRAGSTATS and Patch Analyst (GIS based):

- Area & edge metrics
- Shape metrics
- Contrast metrics

- Aggregation metrics
- Isolation metrics
- Diversity metrics

The above landscape metrics measure two different aspects of landscapes—1) composition or the variety and quantity of patch types without reference to their spatial arrangement 2) configuration or the spatial character and arrangement of patches. These metrics measure characteristics such as distance between patches of the same type and clustering of patches and patch types.

Furthermore, landscape metrics can be computed at three different levels:

i. *Patch level*: Patch-level metrics describe characteristics of individual patches such as size, shape and nearest neighbor distance. Each patch is assigned unique values.

ii. *Class level*: a class is comprised of a set of patches of the same type. Examples of class-level metrics include average patch size and degree of aggregation or clumping computed for all patches within the class. Class-level metrics are considered particularly important for landscape research as they are representative of phenomena such as fragmentation that have important implications for biodiversity and resilience.

iii. *Landscape level*: Metrics at this level compute values across the entire area of interest (patch mosaic) without differentiating between patches and classes (sum or average across all patch types). Typical measurements generated include diversity of patch types, average patch size and degree of clumping.

In addition to describing landscapes at a point in time, Leitao et al (2006) list landscape change and resulting land transformation, habitat loss and fragmentation as key characteristics of landscape ecology research. Understanding landscape change precipitated by anthropogenic disturbances such as urbanization is key to landscape health diagnosis and prognosis. In other words, quantifying, analyzing and monitoring landscape metrics over time is critical for future landscape planning, management and conservation efforts (Leitao et al 2006; Park, 2013; Su et al, 2012; Paudel and Yuan, 2012; Li et al, 2013)

Ueemaa et al (2009) note that the most abundant use of landscape metrics have occurred in studies involving biodiversity, habitat analysis and landuse/landcover change. Research on the utility of landscape metrics in measuring social aspects of landscapes are absent. However, more recently there is an upward trend in studies that examine the relationships between landscape structure and human perception/landscape preference as well as well as landscape regulating functions/ecosystem services (water quality, erosion control) (Uuemaa et al, 2013).

Alberti et al (2003) also propose a conceptual framework in better integrating human systems into ecological research towards creating a framework for sustainable landscapes (Figure3 shown below). Health is one among several other ecological functions that are impacted by landscape pattern.

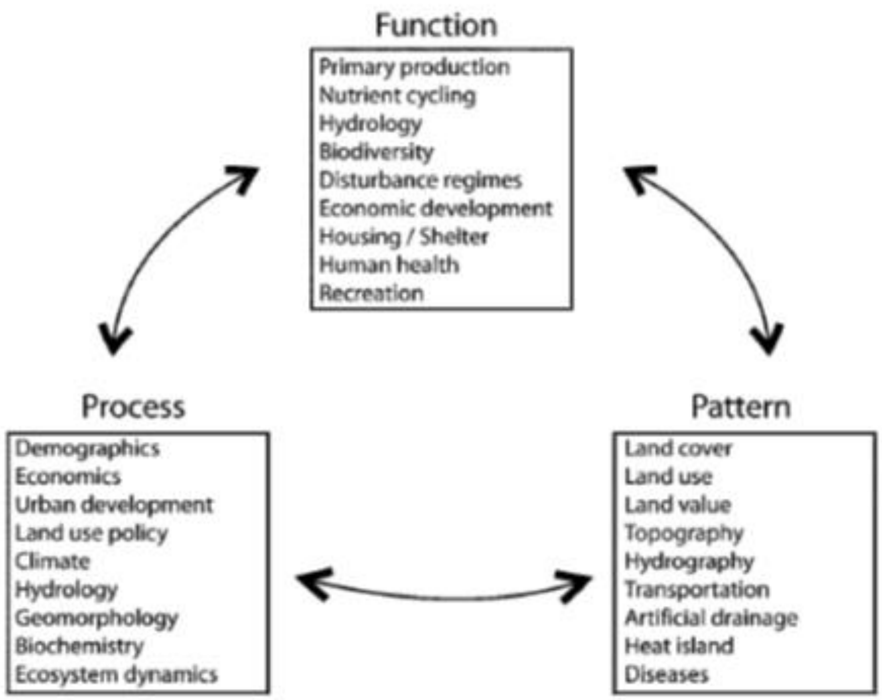


Figure 3. .Conceptual Framework for Sustainable Development (Image Source: Alberti, 2003)

The framework shown above indicates that several deductions and associations can be potentially researched in connection with human health and landscape metrics. For example, contagion metrics (measure of clumpiness of land cover types) has been found to positively correlate with landscape complexity and biodiversity (Uueemaa et al, 2013, Leitao et al, 2006). Complexity has in turn has been found to be positively associated with landscape preference and walkability (Ewing and Handy, 2009; Ewing et al, 2006; Kaplan and Kaplan, 1989). Thus, landscape metrics could build a good (and easily usable tool) foundation for assessing landscape visual quality. This could have associations with higher rates of walkability in communities and could potentially guide the design of green infrastructure. Dean et al (2011) even hypothesize that biodiversity has mental health benefits in making a case for improved land conservation.

Drawing from interdisciplinary approaches based on landscape ecology, population dynamics, urban economics and complex systems science, landscape metrics are used as a method of quantifying the interactions between urban development and natural ecosystems functions/dynamics. Termed “ecological signatures” or “landscape signatures” researchers used landscape metrics to develop typologies of land cover and land uses. Statistical analyses were then performed to understand relationships between these typologies and indicators of ecosystem health such as biodiversity across terrestrial and aquatic habitats. The aim of these studies is to guide urban development patterns and minimize their role as disturbance regimes while preserving the integrity of natural ecosystem functions (Alberti, 2005; Alberti, 2007; Alberti and Marzluff, 2004).

Landscape Patterns and Human Health

A limited number of studies examine the associations between land cover patterns and human health. However, the connections between landcover patterns and zoonotic infectious diseases are relatively more abundant. Within a Landscape Epidemiology approach, Lambin et al (2010) suggest a set of ten propositions to characterize “pathogenic landscapes”. They summarize evidence from several studies on infectious diseases transmitted through zoonotic vectors (diseases such as Encephalitis, Malaria and Lyme disease). A key evidence-based proposition is that disease prevalence and transmission are dependent not only on landscape composition but configuration as well.

For example, vector densities have been found to be directly related to certain types of fragmentation and edge conditions as measured by landscape metrics.

In a study on obesity, physical activity and landscape patterns, Kim (2013) found that more connected landscapes and larger patches of urban forest were negatively associated with BMI. The utility of the study is limited as it examines a sample of 61 Hispanic children in an inner-city neighborhood in Houston, Texas. However, the implications of the study are interesting for this dissertation as it starts to conceptualize a framework that can measure structural and aesthetic qualities of the environment simultaneously (Figure4). This is productive in devising metrics that can potentially capture objective and subjective qualities of the physical environment that support healthy behaviors (physical activity).

	Ecological Criteria	**Health Criteria**	**Proposed Landscape Indices * (acronym)**
Fragmentation	Unfragmented landscape structure	Existence of landscape structure	Number of Patches (NP), Patch Density (PD), Mean Patch Size (MPS), Fragmentation Measurement Index (FMI)
Size	Larger patch size	Size of landscape structure	Total Area (TA), Percentage of Landscape (PLAND), Total Edge (TE)
Shape	Irregular shaped boundaries of patches	Formal or artistic attribute of landscape structure	Landscape Shape Index (LSI), Mean Shape Index (MSI)
Isolation	Closer distance between single patches	-	Mean Nearest Neighbor Distance (MNN)
Connectivity	Connectivity	Connectivity of landscape structure	Patch Cohesion Index (COHESION)

Figure 4. Hypothesized connections between landscape pattern of Green Infrastructure and Health (Source: Kim, 2013)

This section of the proposal summarized several important bodies of literature that contribute to the theoretical and methodological premise of this proposal. In the following section, the literature is synthesized to highlight important themes and gaps. It also summarizes how the proposed research fills the gaps and its contributions to the current state of knowledge.

Synthesis, Research Gaps and Contributions of research

1. Create a more unified and comprehensive approach to measuring the built environment within the socioecological model

The explication of "healthy places" is currently a fragmented and discipline-specific endeavor. The fragmentation occurs along two primary axes. The first type of dichotomy presents itself in the philosophical approach to defining health and the man/nature duality that gets emphasized as a result. The second rift is evident in the methods and metrics that are used to evaluate the impacts of the environmental envelope on human health. While most approaches use reductionist methods, multilevel modeling and other complex systems approaches are gradually gaining momentum.

Several disciplines represent different aspects of the socioecological model of public health. For example, landscape epidemiology explores topological relationships of the natural environment that contribute to disease risk and causation. Urban planning looks at accessibility to services and measurements of urban form as they relate to human health (anthropocentric or human-centered). Meanwhile, there is a growing consensus that human well-being is interdependent on the health of natural systems.

Ecological sciences have a different methodological approach to measuring health of ecosystems that often exclude human factors and well-being (biocentric or nature-centered). This dissertation makes two important theoretical contributions with regard to the above—1) it uses a method typically employed in evaluating ecosystem health (landscape pattern analysis/landscape metrics) to establish connections between landscape patterns and human health. For example, we know that certain values of landscape metrics such as patch-size, connectivity, etc. are positively associated with ecosystem health indicators such as biodiversity. While the dissertation does not explicitly consider ecosystem indicators in the analysis, it establishes a baseline of associations between landscape metrics and human health. By establishing thresholds of landscape patterns associated with human health, a balanced approach to healthy land use planning can be developed in the long term. This incorporates the best interests of human and other species. 2) It creates a more holistic expansion of the socioecological model of public health by using methodological and operational constructs from different disciplines (inclusion of social, physical and natural patterns).

Social indicators such as socioeconomic status (SES) and demographics are rarely used in assessing natural ecosystem health. It is interesting to note that concepts of patterns analysis are also employed in social epidemiology. For example, patterns of residential segregation are considered important predictors of regional health outcomes (infant and adult mortality rates). Furthermore, residential segregation is characterized as a multi-dimensional construct consisting of unevenness, isolation, clustering, centralization and concentration. These constructs contain parallels with landscape metrics described above. However, most studies emphasize just one dimension at the metro scale, and neighborhood deprivation is often considered the operating process between pattern and outcomes (Acevedo-Garcia, 2003). This dissertation will incorporate it as an important descriptor of patches instead of just a control variable. In summary, pattern analysis is applied to social determinants as well.

2. Understanding the entire mosaic of land uses and their contribution to health outcomes enhances existing compositional measurements such as sprawl

The landscape is a composition (mosaic) of individual elements and the interrelationships between them (configuration/juxtapositions). Current approaches to the connection between urban form and health is restricted to the study of various sprawl attributes/individual land uses and health outcomes/determinants. Further, they do not investigate the relationship of land use as a landscape. Sprawl is often characterized by density, compactness and measures of physical attributes. There is also little consensus regarding which land uses contribute the most to health outcomes. Primary land uses explored include housing and employment but can vary greatly. Again, these are operationalized through quantity rather than spatial arrangement. This approach does not take into account interactions between these uses. This research seeks to answer the question “How is mix, spatial distribution and proportions of land uses associated with health outcomes (among several other questions)?

3. Use landscape pattern analysis (landscape metrics) as a new, more nuanced methodological approach to guide healthy land use planning

Landscape Pattern analysis is a relatively new, growing methodological field in Landscape Ecology. It posits that landscape patterns are indicative of underlying processes that contribute to associated outcomes. Landscape Pattern Analysis is starting to be used in Urban Planning to understand sustainable urban form. However, these studies utilize satellite imagery where land cover patterns are used for broad approximations of land use. The emphasis is on understanding ecosystem impacts. This research uses actual land use data relevant to human uses to broaden the scope of this research to human health. Further, it explores how changes in land use patterns over time can also impact health. This approach is particularly useful in informing land use policy and development.

4. Understand the different scales at which landscape patterns matter

Regional patterns influence local processes. However, we do not how regional land use patterns influence neighborhood patterns and health outcomes. This research answers the question "At what scale do landuse patterns impact health outcomes?" The importance of land use mix is often utilized in walkability studies at the neighborhood level. However, these are aggregated indices often combined in subjective weights. This research provides objective calibration of how such indices can be further refined. Furthermore, we know that land use patterns impact health at different scales. While a green infrastructure approach at a regional scale might have implications for air and water quality, land use mix at the neighborhood scale impacts obesity and walkability. This research provides a comprehensive approach to begin prioritizing land use decisions at different scales in the interest of creating healthy places and regions.

CHAPTER III

RESEARCH FRAMEWORK

Methodological Framework

This dissertation makes the following important contributions to the built environment and health literature:

Methodologically, it introduces a new way of measuring urban form that incorporates compositional and configurational characteristics. This method, currently employed in Landscape Ecology to measure land cover patterns is easily adaptable to studying land use patterns. It is also a highly scalable method in that patterns can be generated at any scale, allowing us to simultaneously measure objective and subjective characteristics of urban form. Several ways of utilizing this methodology are envisioned through this dissertation and provide an agenda for future research. When analyzed at the regional scale, landscape metrics can be used to characterize larger ecosystem processes. When analyzed at the neighborhood scale, they can provide insights into subjective factors such as landscape complexity and visual preferences that potentially promote healthy behaviors.

The approach described above provides an important bridge between current fragmentation in regional and neighborhood level studies. The dissertation tests direct associations between urban form and health outcomes. The rationale is to examine whether this more nuanced method of measuring urban form has direct explanatory potential for benchmarking land use and health associations at multiple scales. The pathways between urban form and health outcomes are numerous. Since this research presents a new methodology of measuring urban form, the primary attempt is to capture a cumulative effect on health. This dissertation presents a first-level investigation between land use patterns and systemic measures of population health (mortality). It also examines the utility of existing data structures in providing a parsimonious approach to measuring determinants of regional health. Unpacking the pathways is important nonetheless, but beyond the scope of this immediate undertaking. The mediated connections between urban form, health, determinants and health outcomes is to be followed up in further research.

Theoretically, it incorporates multidisciplinary definitions of healthy places to create a consolidated approach. The most prominent approach to measuring land use patterns and health is currently the urban sprawl and health study conducted by Ewing and Hamidi (2014), which is an update on the same study conducted in 2003. While the study is useful with regard to characterizing land use composition, it provides little information with respect to land use configuration. In other words, counties would get the same sprawl scores based on composition even if their configurations were different (Figure 5 and table.1). However, a closer examination of select landscape metrics (county scale) in both counties reveal differences in the configuration of different land use typologies. While certain metrics such as Shannon Diversity Index are similar (parallel to sprawl metrics), patch level landscape metrics exhibit greater variation (Figure6-8). A more detailed explanation follows after the figures (on pg.34).

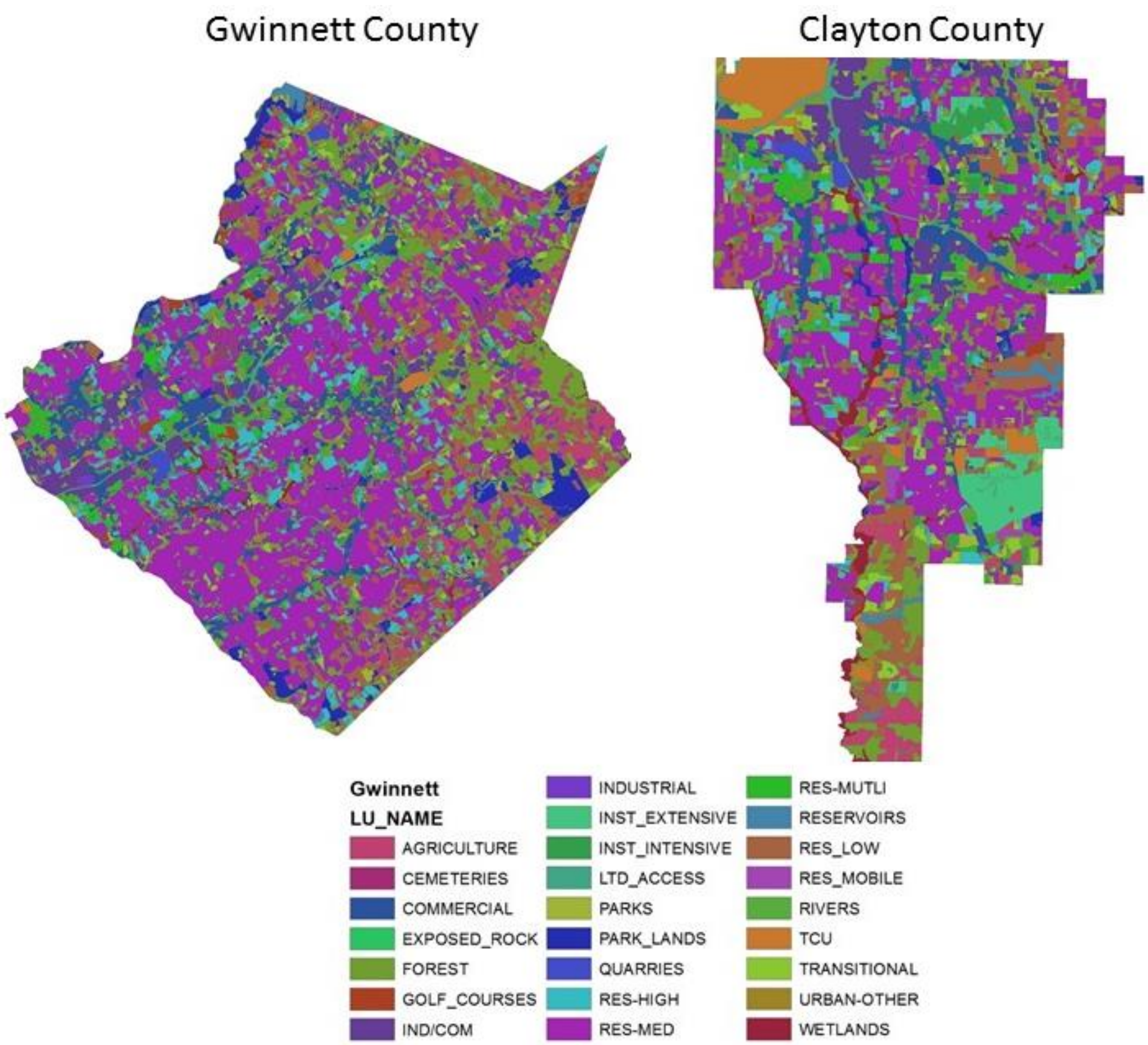


Figure 5. Comparison of Gwinnett and Clayton County land use patterns

Table 1. Sprawl Metrics for Gwinnett and Clayton counties as computed across four domains. This methodology and corresponding metrics are based on the most recent update of a previous nationwide sprawl study (Ewing and Hamidi, 2014). Data source: http://gis.cancer.gov/tools/urban-sprawl/

County	Density factor	Mix factor	Centering factor	Street factor	Composite index 2010
Gwinnett	106.36	111.9	88.7	89.68	98.95
Clayton	106.35	106.2	84.62	98.1	98.49

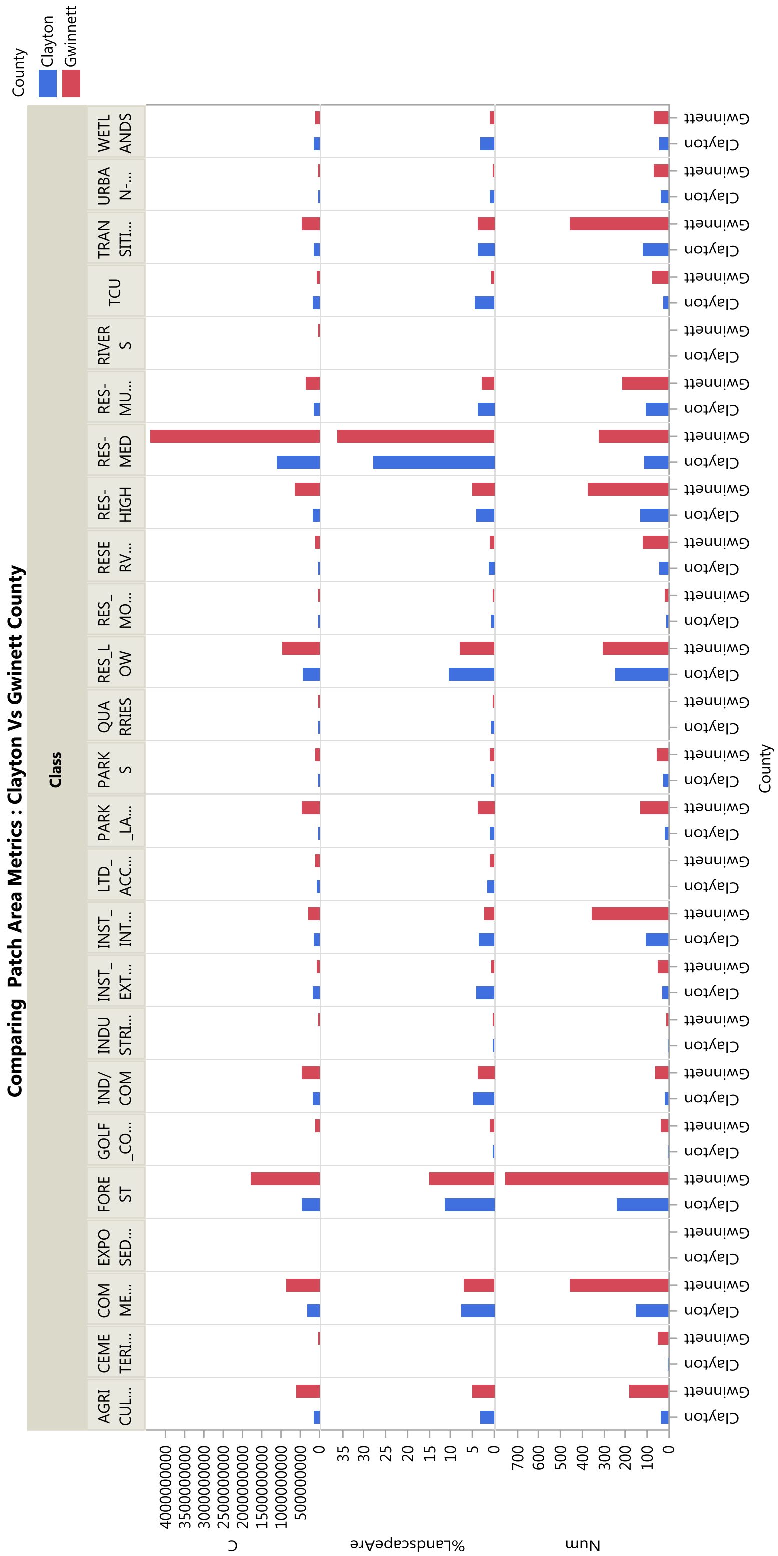


Figure 6. Gwinnet & Clayton Counties: A comparison of patch area metrics of constituent land uses

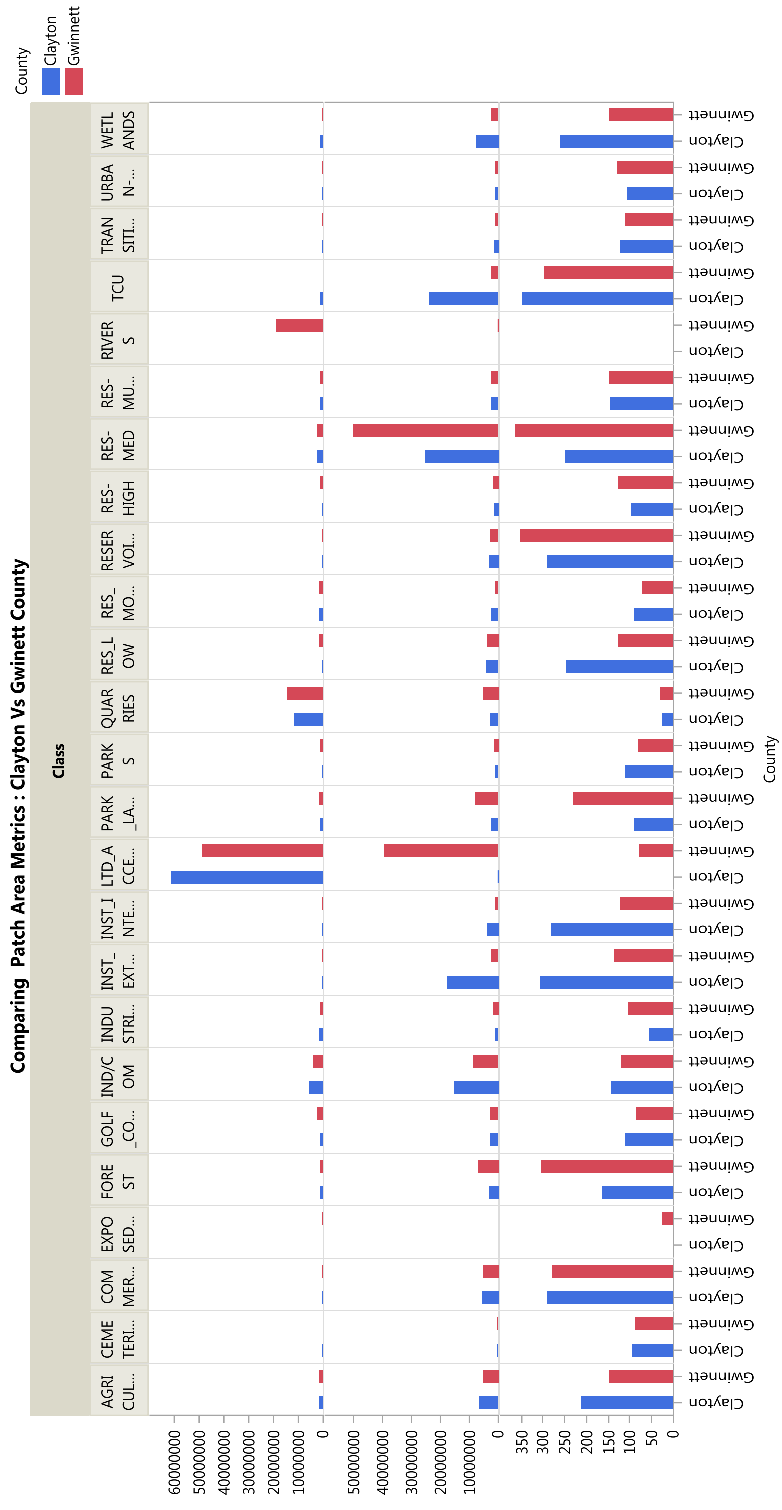


Figure 7. Clayton & Gwinnett: A comparison of patch diversity & size metrics

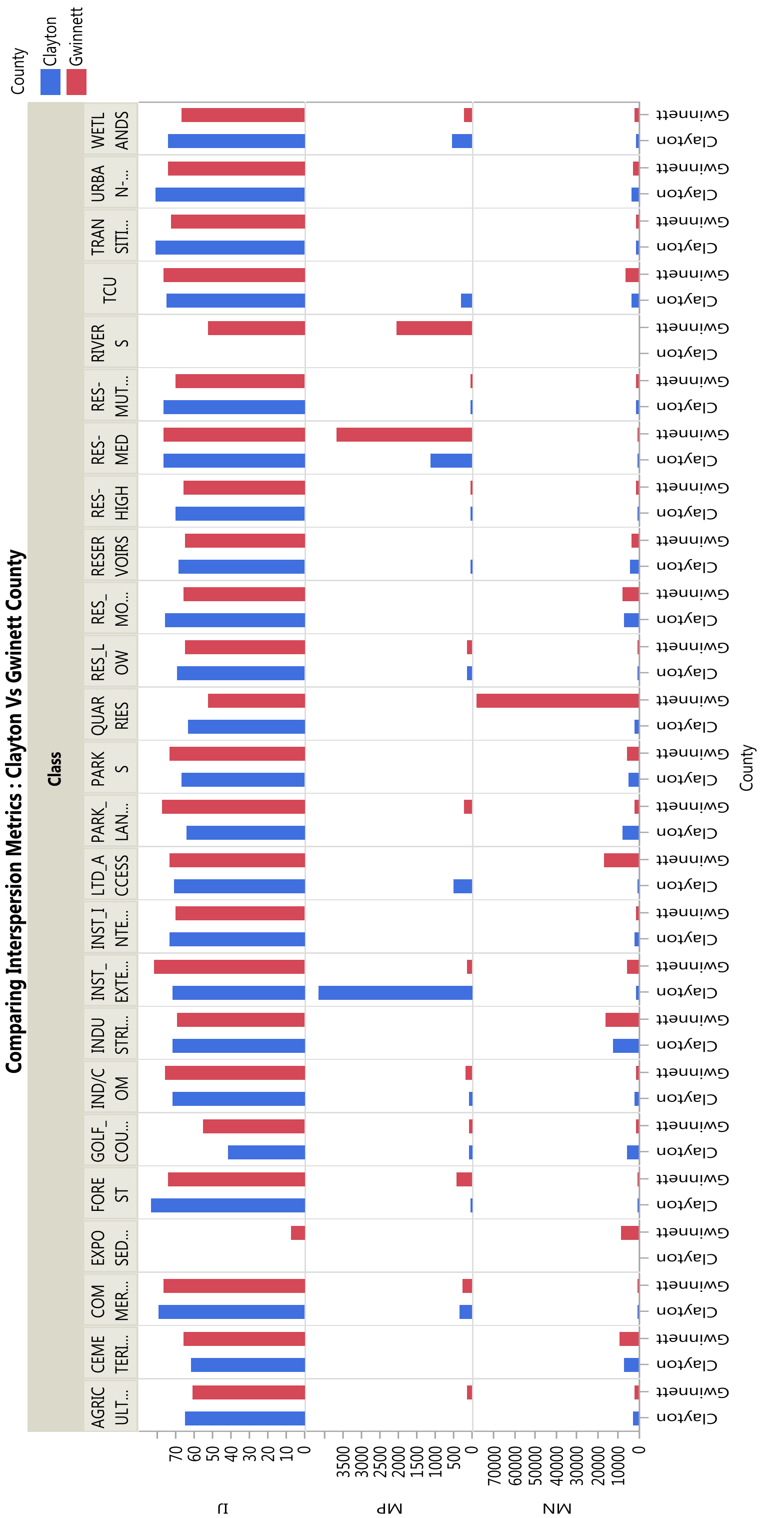


Figure 8. Clayton & Gwinnett: A comparison of diversity and interspersion metrics. These metrics measure the degree of adjacency, isolation and fragmentation of patches

The graphs in Figures 6-8 provide a rationale for the research approach in this dissertation. Figure5 and table 1 indicate that Gwinnett and Clayton counties fare similarly with regard to the various measures of sprawl. However, the bar graphs characterizing the different values of their landscape configuration show greater variation. For example, when we look at Commercial land use between the two counties, Gwinnett has a significantly higher number of patches and total class area (Figure6), marginally lower PSCoV (Figure7) and Mean Proximity Index (Figure8). These potentially indicate that Clayton may have larger patches of Commercial land uses in closer proximity. The bar graphs also represent the shape, diversity and spatial distribution of the various land use typologies in the two counties.   In other words, this dissertation makes the claim that disaggregate measures such as landscape metrics that capture spatial arrangement and distribution of land use might be an improvement over more commonly used measures of urban form (such as sprawl). This variability in landscape metrics is also captured in the hierarchical clustering diagrams (Figure 12 & 13).The variation in configuration metrics along with variation in social and health metrics is further investigated.

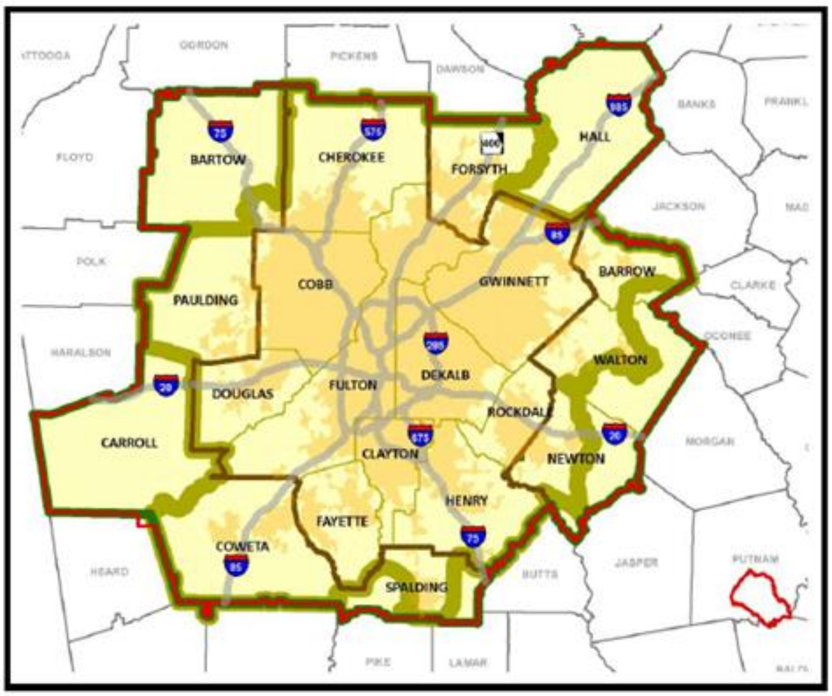
Geographical location of study area

The study area for this dissertation is the 21 county region under the jurisdiction of the Atlanta Regional Commission (ARC) which is also the regional Metropolitan Planning Organization (MPO) for the Atlanta Metropolitan area (Figure9 & 10). Atlanta has a history of attempting to incorporate health into urban planning. Back in 1997, the SMARTRAQ study was a pioneering effort in understanding the relationships between urban form, travel behavior and health outcomes. More recently, the Atlanta Regional Commission (ARC) is actively trying to incorporate health metrics into its regional land use and transportation planning (PLAN 2040) and decision-making endeavors. This is evidenced through the Health Impact Assessment (HIA) of the PLAN 2040, one among many other nationally prominent HIAs conducted by the Center for Quality Growth and Regional Development (CQGRD) at Georgia Institute of Technology. In addition, there are several synergistic relationships underway between the CQGRD, the Centers for Disease Control (CDC) and Georgia Department of Public Health (Ga DPH) in furthering built environment and health research. This research is envisioned as an important contribution to this regional research agenda.

Defining the landscape at multiple scales- the Atlanta metro area

Garigal et al (2012) state that "landscape is not necessarily defined by its size; rather, it is defined by an interacting mosaic of patches relevant to the phenomenon under consideration (at any scale). It is incumbent upon the investigator or manager to define landscape in an appropriate manner. The essential first step in any landscape-level research or management endeavor is to define the landscape, and this is of course a prerequisite to quantifying landscape patterns."

Defining health at the regional level is a new endeavor. For the purpose of this dissertation, the Atlanta Metropolitan Area will be considered for analysis. This represents a functional region that the ARC oversees as its planning jurisdiction. From an organism-centered perspective, it represents a functional area from a human perspective, composed of counties and census tracts (neighborhoods). It represents an area of conscious, physical and political decision-making that directly impacts human health. Land use and comprehensive planning usually occurs at these two levels as well. For analytical purposes (multilevel modeling), two geographical levels are specified in the definition of landscape. The tract level variables are specified as level 1 variables (lowest unit of analysis) and county level variables are specified as level 2 variables. The cumulative landscape is thus comprised of 1000 census tracts nested within 21 counties (Figure10). Landscape metrics are then generated at the landscape and class level for both geographical units.



Color				
Boundary Name	Regional Commission (RC)	Metropolitan Planning Organization (MPO)	Ozone Non-Attainment Area (8 hour standard)	Particulate Matter (PM 2.5) Non-Attainment Area
Number of Counties	10 counties	All of 13 counties; parts of 5 counties	20 counties	All of 20 counties; parts of 2 counties

Figure 9. ARC jurisdictional boundaries (Image source: Atlanta Regional Commission, 2016)

Spatial Level	Geographical Unit	Landscape boundary
1	Census Tract	Census Tract boundary
2	County	County boundary

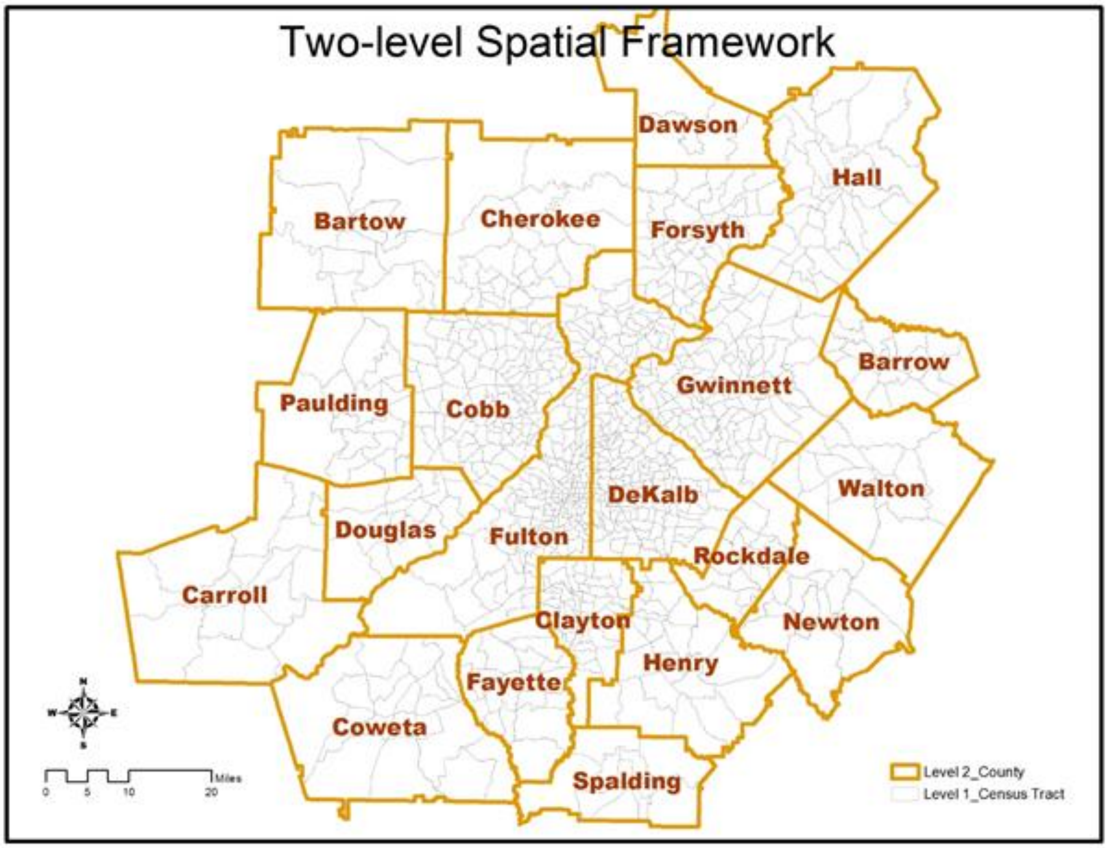


Figure 10. The two-level spatial framework for the Atlanta Metropolitan Region with tracts nested within counties

Conceptual Framework

This dissertation characterizes the landscape as a cumulative phenomenon of physical attributes (land use/land cover patterns) as well as social attributes (social patterns). Social attributes are known mediators of the relationship between built environment and health (Lathey, et al, 2009; Bodea et al, 2009). A widely acknowledged challenge in Landscape Pattern Analysis is understanding the underlying processes that bridge the relationship between landscape patterns and quality of ecosystems. There is abundant research on land use and its impact on health determinants such as air quality, transportation behavior and physical activity. Urban form is usually measured as sprawl (pattern variables) and its impact is tested on health determinants (process variables) with the objective of explaining health outcomes (Figure 11).

A new, more nuanced method of measuring urban form and its relationship to health outcomes is investigated in this research. Landscape Pattern Analysis enhances our understanding of urban form by describing configuration in addition to composition. Therefore, the methodology seeks to first benchmark connections between patterns and health outcomes, while controlling for social patterns. Health effects are the directly measured health outcomes in the form of mortality, morbidity and disease rates. Health determinants are defined as the mechanisms that support healthy behaviors and outcomes. For example, certain land use patterns might support better access to parks, in turn supporting higher levels of physical activity. While health determinants do not "determine" a health outcome (implying a causal relationship) they might increase or decrease the risk for disease (associative relationship).

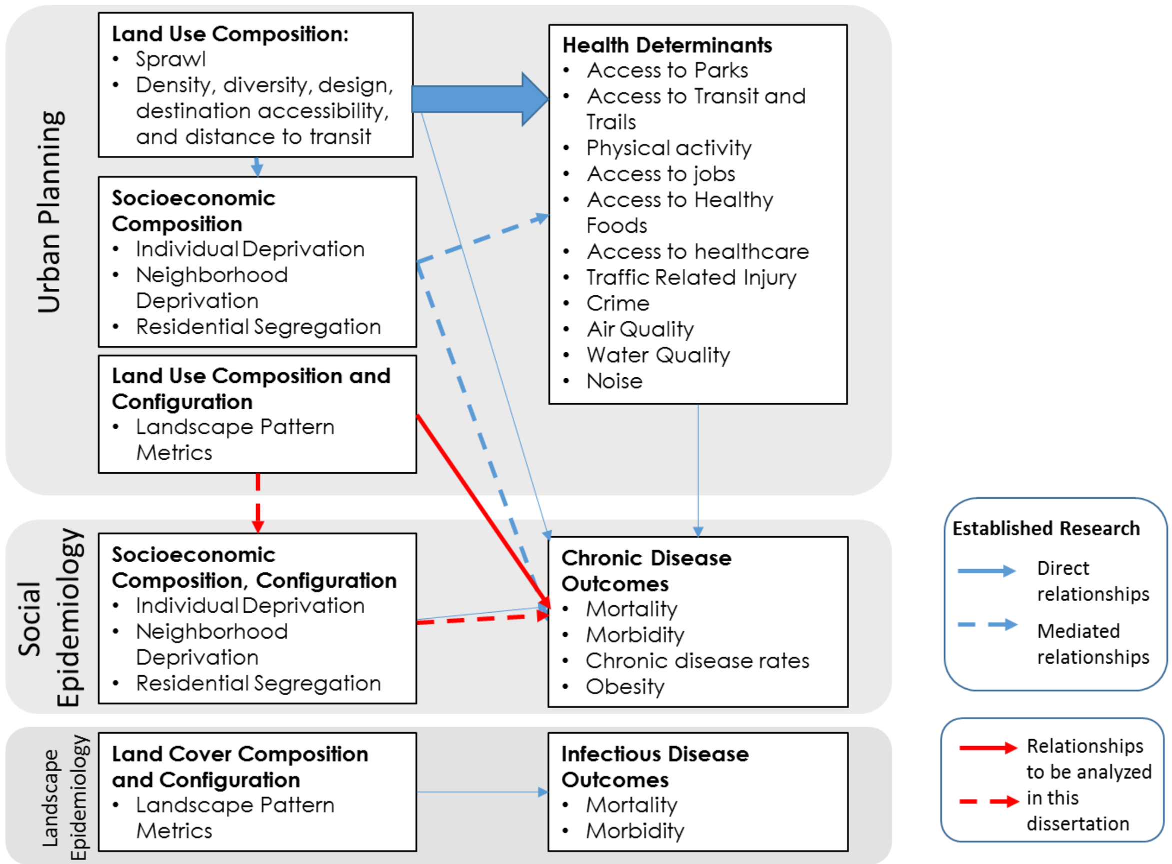


Figure 11. This diagram illustrates the various discipline specific pathways analyzed in built environment and health research. The red arrows show the hypothesized relationship between landscape patterns and health effects as investigated in this dissertation.

Research Questions and Hypothesis

Three primary research questions are investigated in this dissertation:

1. ***Are landscape patterns important determinants of human health?***

- *Are certain land uses and their patterns significantly correlated with health outcomes as measured by mortality rates?*
- *Do social patterns have a stronger association with mortality rates compared to landscape patterns or is there an interactive effect?*

- *Do social patterns show consistency with landscape patterns or do they have independent roles in influencing health outcomes?*

2. ***How is land use mix and spatial distribution (composition and configuration) of landscape components (land use and socioeconomics) associated with health outcomes?***
3. ***At what scale do landscape patterns significantly impact human health outcomes?***

The primary hypothesis used to investigate the above questions are stated below:

H_a: Landscape patterns characterized by high heterogeneity and juxtapositions of complementary land uses and social characteristics are inversely associated with mortality rates.

H_o: Landscape patterns characterized by high heterogeneity and juxtapositions of complementary land uses and social characteristics have no association with mortality rates.

Land use heterogeneity (characterized as land use mix) is an established correlate of physical activity and other social phenomena that influence health outcomes. However, in current methodological approaches, this is measured as a compositional construct with little differentiation between the spatial distribution of various land use types and their complementary adjacencies (Ewing et al, 2014; Hess et al, 2001). **For the purpose of this dissertation, the landscape is theoretically defined as a composite of land use and social characteristics, a combined construct of patterns and process.** Landscape pattern metrics provide a more nuanced and contextual approach to measuring land use/social heterogeneity. They provide a better understanding of complementary adjacencies and distributions through measures such as the "interspersion juxtaposition index (IJI)". The IJI is a unique metric that explicitly measures spatial configuration of patch types (neighborhood relationships of patches).The metric measures the extent to which a patch shares its edge with another patch type (patch adjacency). Low values imply landscapes where patches are clumped or are bordering only a few other classes. High values denote landscapes where patches are equally adjacent to each other.

Analytical Framework

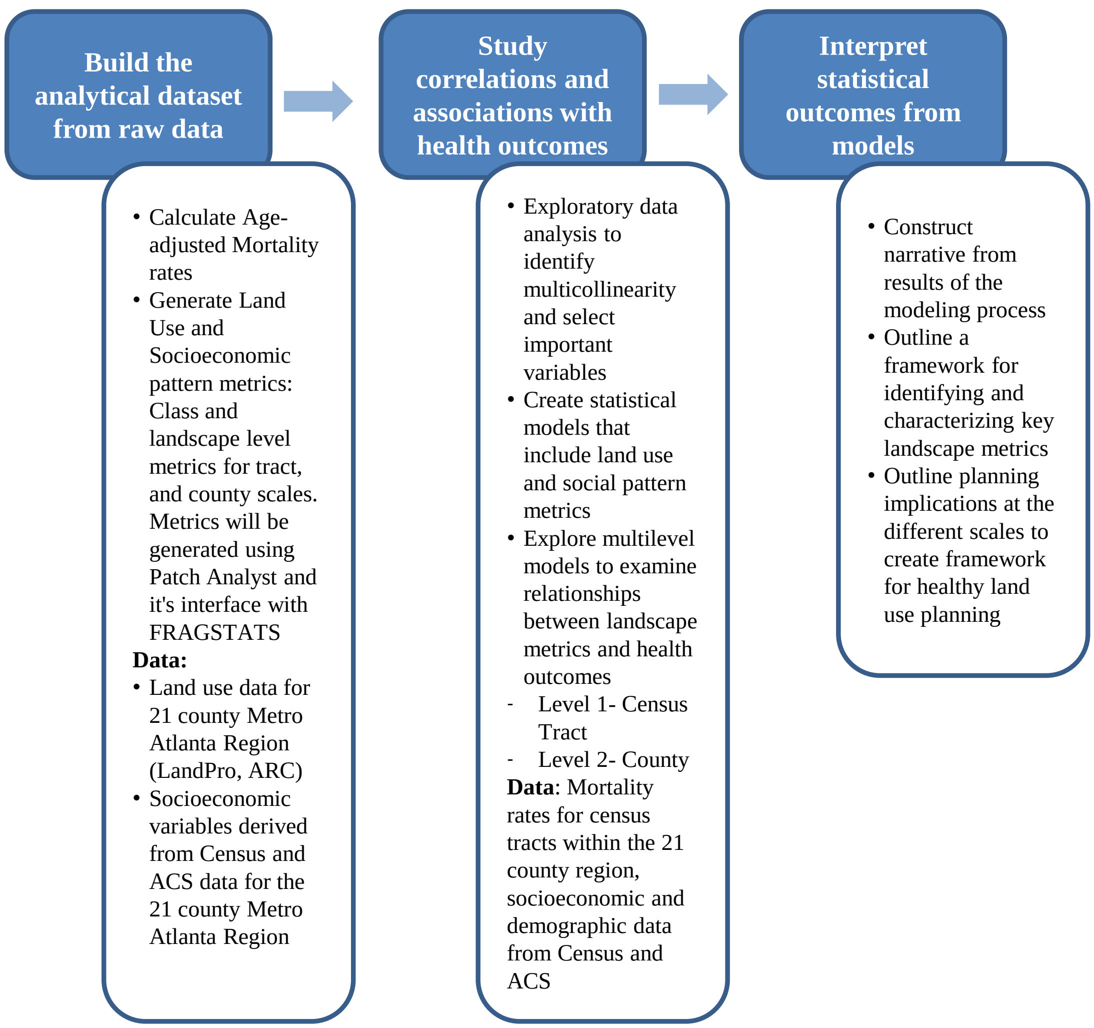


Figure 12. Steps in the Analytical Process

Methodology

This section provides an overview of the analytical procedures undertaken in this dissertation and the rationale that guided that process. A more detailed discussion of the specific analytical steps and the incremental process used to create the final variables for analysis are provided in Chapter IV. An overview of the entire process is also illustrated in Figure 17.

Stage 1: Assembling the Dataset

Step 1: Calculate age-adjusted mortality rates for census tracts

Raw mortality data (total counts) stratified by age, gender, race and cause of death was provided by the Ga Department of Public Health from 2002-2011(cumulative 10 year counts). Age-adjusted rates for select diseases were calculated to better account for population distributions that can bias mortality rates. For the state of Ga, mortality rates show a reasonably close relationship with morbidity rates (as measured by hospital discharge data) and so will be considered an indicator of disease prevalence. A correlation analysis of urban counties (including those in the study area) yielded a result of .71 for all-cause mortality and morbidity rates (age-adjusted) and a result of .6 for cancer mortality and morbidity (data source: Ga DPH OASIS).

A more detailed explanation of the selection of disease types and their conversion into a binary dependent variable is provided in Chapter IV.

Step 2: Constructing the dependent variable

Univariate techniques such as histograms and box-plots were generated to gain a preliminary understanding of distributions, spread and outliers of the age-adjusted mortality rates calculated in Step 1. Correlation matrices were generated to detect multicollinearity. There were no universal pattern of correlations across all disease types (Figure 13.) Therefore the decision was made to analyze health conditions separately. This also helped reveal different environmental mechanisms and relationships relevant for the different disease types. However, it is interesting to note certain subtle trends in the correlation diagram. Since the correlations are clustered, respiratory illnesses (Lung Cancer, COPD and Pneumonia) appear in one group while diabetes and heart disease have a stronger correlation as well (0.5).

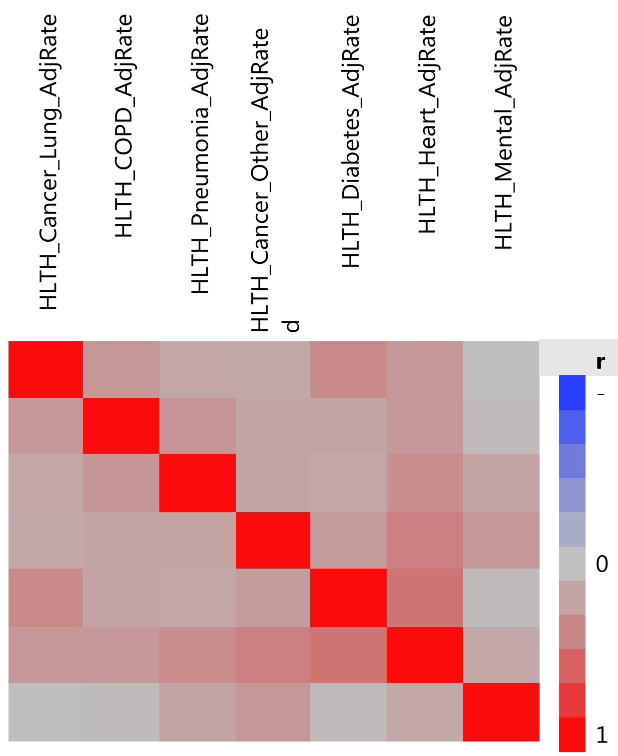


Figure 13.Age-Adjusted Rates Clustered by Correlations. Stronger correlations are represented by darker shades of red.

The final goal is to create the categorical dependent variable for this research. Advanced clustering methods will be used to further explore data patterns. The purpose of clustering techniques is to find optimal groupings for which cases are similar. Spatial clustering was explored in GeoDa through an analysis of Local Indicators of Spatial Association (LISA). The results of the clustering analysis were mapped and threshold for hi/low mortality rates were selected from the clustering analysis. This threshold was then used to create the binary, categorical dependent variable. Chapter IV presents a detailed explanation of spatial clustering, the results of the analysis and subsequent derivation of the dependent variable.

Step 3: Constructing the Independent variables

Two sets of independent variables were used in this analysis— 1) Land use Pattern metrics 2) Neighborhood Deprivation Index. Land use patterns for selected land uses were generated at the census tract scale. Correlation studies led to the decision of creating two types of land use indices using Principal Components Analysis. The first index was an aggregated index combining all metrics for each land use. The second index was a disaggregated index where groups were created from the specific domain that the metric measures. Each land use has three such indices based on Geometry, Shape and Interspersion.

A single Neighborhood Deprivation Index was calculated for each census tract . The variables and methodology used to create this index were derived from a rigorous literature review. By definition, census tracts are intended to represent fairly homogenous areas with respect to socioeconomic characteristics. A similar approach to describing demographic trends (segregation and fragmentation in race and income) has been explored by Crews and Peralvo (2008).

Stage 2: Data Analysis

Both Exploratory and Confirmatory Analysis were performed on the dataset with the following objectives:

Exploratory Analysis: Hierarchical Clustering was used to visualize any readily discernible patterns or groupings in the data and provide insights that could be carried over into the confirmatory models.

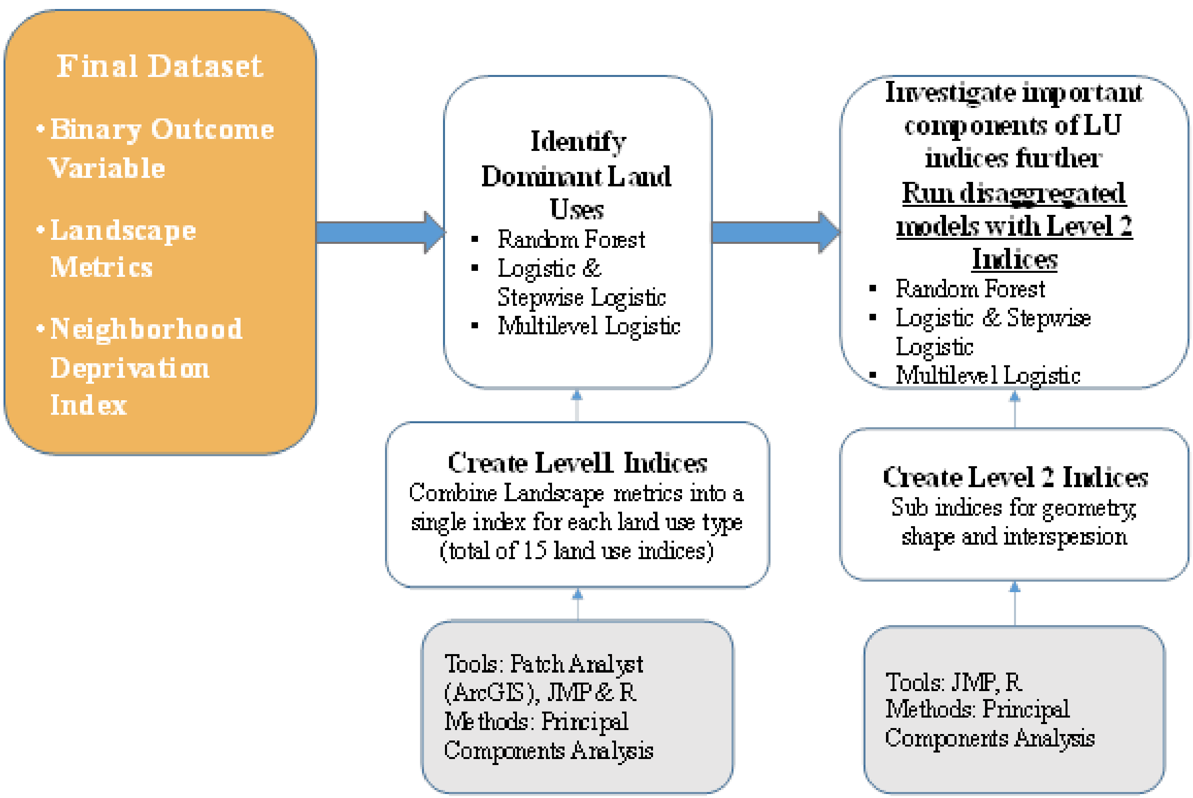


Figure 14. Overview of the Confirmatory Modeling Process

Confirmatory Analysis: The aim of the Confirmatory Analysis was to identify independent variables that were statistically significant and understand the magnitude of their contributions to the health outcomes of interest. A series of modeling methods were implemented to incrementally detect significant effects. These models include both traditional classification methods as well as more recent data mining approaches. The data mining approaches pick up nonlinear relationships that might otherwise be missed by traditional approaches. Another analytical objective was to look for consistency among modeling results. The models were implemented using base R and packages including 'stats', 'caret', 'Random Forest', 'lme4' and their dependencies. The caret (short for classification and regression training) package in R was used for creating training and validation datasets. This is the first step in developing a robust model by randomly assigning the dataset into a training and validation portion. This is done in caret, assigning an industry-standard 75% to train the model and use the remaining 25% to see how well the model fits data that was not in the training sample. Caret also adds additional robustness by using repeated cross-validation in training itself, and repeats the whole process multiple times (in our case 5-fold cross validation with 10 repeats) to remove any bias that may occur due to random sampling. This is unique to this package, and is one of the reasons the models tend to select the right variables and predict robustly in practice.

As explained in Figure15 below, the training and validation sets are run through the four different modeling procedures and the entire process is repeated five times. The Random Forest, an ensemble learning method for classification, was applied to the dataset to extract variable importance. Regular, Stepwise and Multilevel Logistic

Regression was performed on the full dataset using all the Level 1 land use indices and neighborhood deprivation index.

Significant variables from the Level 1 analysis were carried through to the Level 2 analysis. Level 2 indices (disaggregated) for each of the significant level 1 land use indices were used along with the Neighborhood deprivation Index and the entire process shown in Figure13 was repeated. This stringent methodology was employed to ensure agreement within and across model types and report results with reasonable confidence.

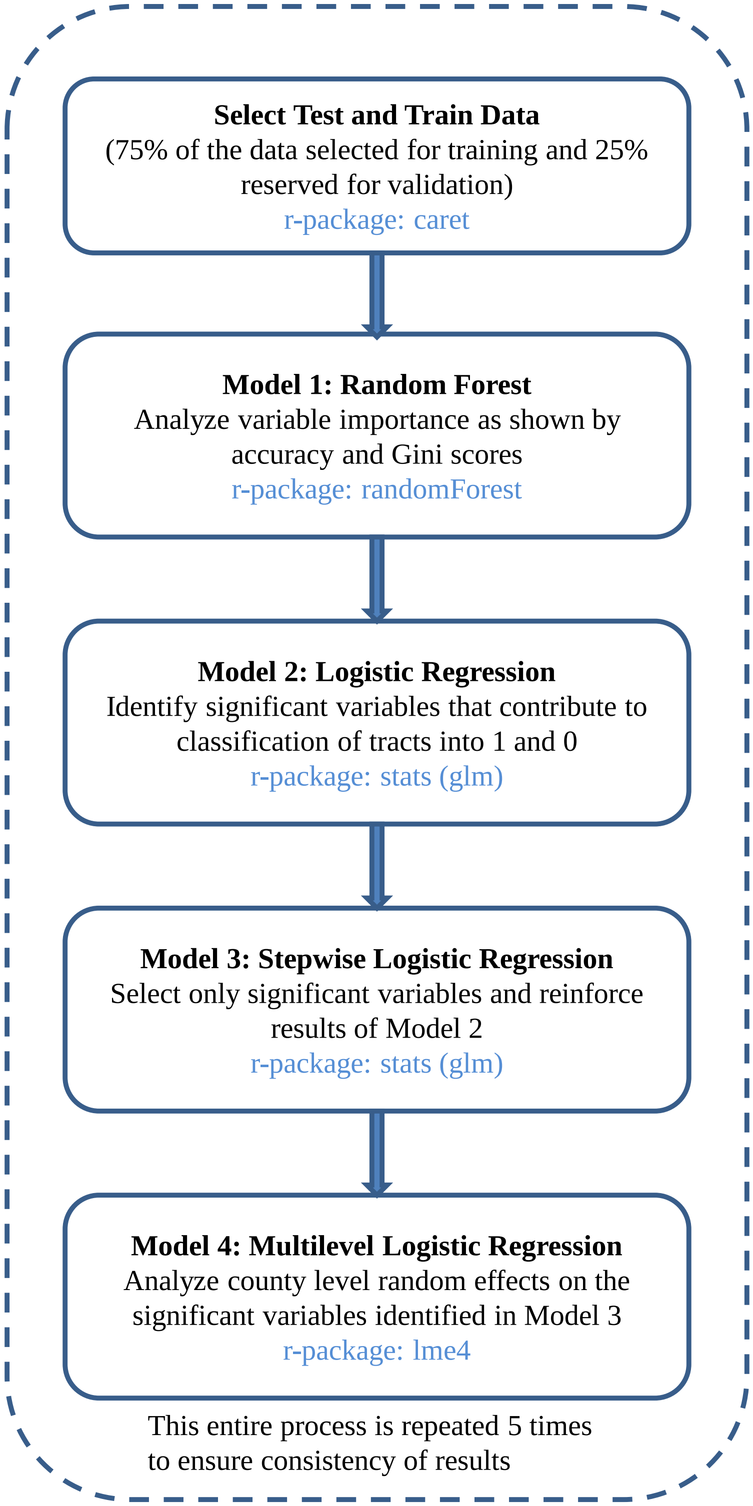


Figure 15. Flow of the modeling process

Step 1: Exploratory Data Analysis (EDA)

Two types of EDA methods are used, namely, **Hierarchical Clustering and Random Forests**. While Hierarchical Clustering looks for discernible patterns that can guide model development, Random Forests aid in selecting important variables.

Hierarchical Clustering: The dissertation deals with a largely "unsupervised" problem as it reveals undiscovered patterns in the data that have not been explored before. Clustering is a particularly powerful exploratory technique to partition a dataset into subgroups based on its underlying attributes. This technique helps discover groupings of similar land use and SES attributes (signatures) that express themselves collectively in describing "healthy" or "unhealthy" census tracts or counties. In particular, Hierarchical Clustering offers two advantages over other clustering methods (such as K-means). It does not require a pre-defined number of clusters and it produces a visually informative, tree-based representation (Dendrogram) of the patterns (Figure16, Figure17). This is a preliminary tool that allows detection of important variables, relationships and patterns pertaining to healthy tracts/counties and other explanatory variables. However, it does not point to the magnitude of relationships. Dendrograms were created for tracts and counties within the study area. They are "hierarchical" as clusters are built bottom-up or agglomeratively, where the lowest level represents groups with the most similar observations, gradually building up more coarse groups (James et al, 2013). This technique is particularly useful to describe county level characteristics where there might be insufficient sample size to conduct traditional regression. The technique is demonstrated at the county level in Figure16-17.

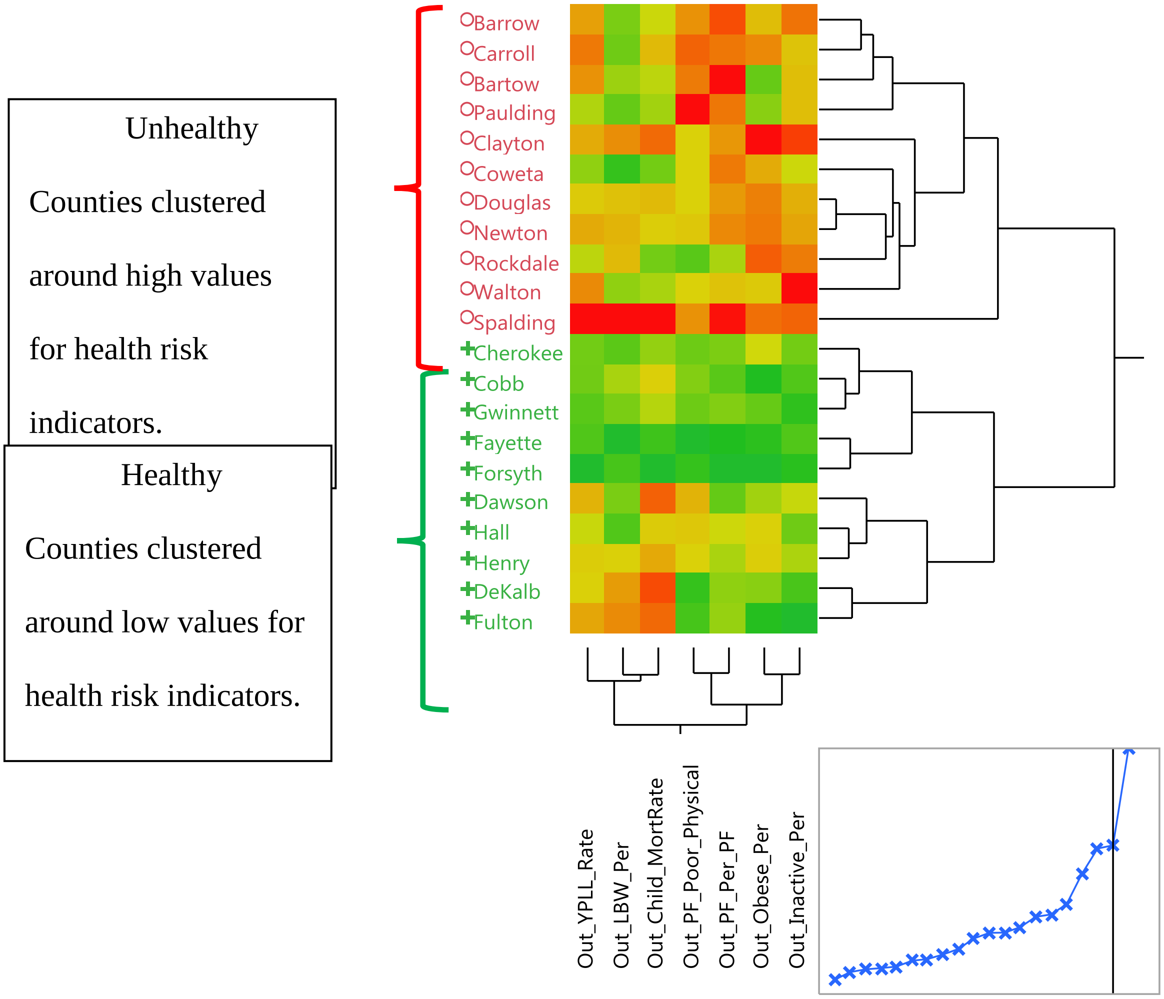


Figure 16. Dendrogram of county clusters based on health outcomes data.

Example of counties clustered around select health indicators listed along the x-axis (Data Source: County Health Rankings, 2010). Two prominent clusters are indicated by counties indexed as green and red. This is a useful method to classify counties into "healthy" (green) and "unhealthy" (red). The counties are assigned to a cluster based on their similarity to other members of the cluster. In this case, the closer two counties are on the Dendrogram, the more similar their health characteristics. The above checker-board can be seen as a matrix where each square represents the value of a certain health characteristic for its corresponding county. The values of the health variables are represented as a spectrum from green (low) to red (high). Note that Gwinnett and Clayton appear in entirely different clusters.

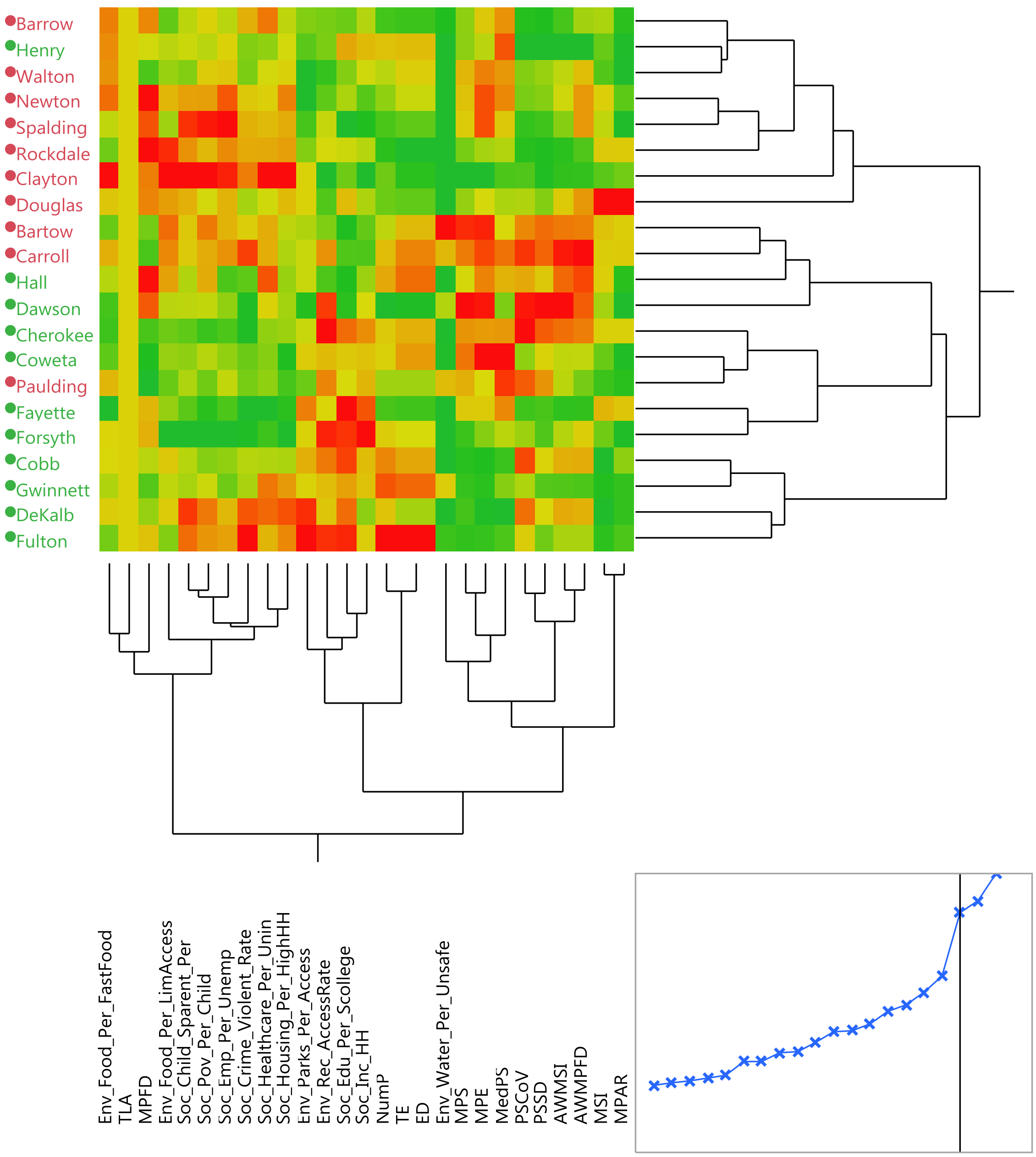


Figure 17. Dendrogram showing clustering of counties, health determinants, land use metrics and SES indicators.

The vertical axis on the left shows the patterning of healthy (green) and unhealthy (red) counties in relationship to landscape and social patterns. When plotted against land metrics and social variables, the clustering of healthy and unhealthy stays fairly intact compared to the previous Dendrogram (Figure 17). This is only an example demonstrated through County Health Rankings data and does not represent the actual data that will be used in the dissertation. Again, Gwinnet and Clayton counties emerge in different clusters. This supports the hypothesis for further investigation that landscape metrics might better capture built environment variability for health research.

Random Forest: Classification using the Random Forest technique is also applied to the data. While used here as a sophisticated exploratory technique, it starts to segway into confirmatory analysis. The random forest algorithm is a powerful, ensemble learning, classification and regression method and has been shown to outperform other well-known classification algorithms. The approach, combines several randomized decision trees and aggregates their predictions by averaging, and has shown excellent performance in settings where there a large number of predictors coupled with potential nonlinear effects. The analysis returns measures of variable importance and thus can inform variable selection for further modeling (Biau and Scornet, 2015; Fern´andez-Delgado et al, 2014).

Step 2: Confirmatory Data Analysis (CDA)

CDA techniques were utilized to accomplish the following objectives:

- Variable selection- Selection of predictors based on statistical significance (step-wise logistic regression). Multiple techniques will be used to confirm consistency of selected predictors
- Computing the magnitude of contributions of predictors towards “healthy” or “unhealthy” status
- Using multilevel modeling techniques to answer the question “At what scale do landscape patterns matter for human health?”

The dependent variable used in this research is categorical, taking on the values of 0 (healthy) or 1 (unhealthy). Thus logistic regression analysis is used to compute the probability that a given tract is healthy or unhealthy based on its landscape signature (land use and SES patterns).

Based on the EDA (data exploration and reduction), a series of logistic regression models were run to answer the research questions:

Logistic Regression: The simplest form of the ordinary logistic regression equation is as follows: [James et al., Page 135]. The outcome of interest is the response probability,$p_i(\boldsymbol{X})$, which is the probability that census tract ‘i’ will be classified in the unhealthy category (Value = 1).

$$log\left(\frac{p_i(\boldsymbol{X})}{1 - p_i(\boldsymbol{X})}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

Where $\boldsymbol{X_i} = (X_1, \ldots, X_k)$ are ‘k’ predictors for census tract ‘i’ that include both land use metrics as well as socioeconomic variables. This equation can also be written in the form,

$$p_i(\boldsymbol{X}) = \frac{e^{(\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\cdots+\beta_k X_{ki})}}{1+e^{(\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\cdots+\beta_k X_{ki})}}$$

Multilevel Logistic Regression: In multilevel logistic regression, the goal is to take into account the effect of hierarchies within the model, where the "lower" level variables ('i') are measured at the census tract level, and these measurements are nested within "higher" level variables ('j') measured at the county level. The model is similar to standard logistic regression, with additional terms to account for variations in the model intercept due to county-specific effects (i.e. random intercept model with fixed slopes). The model is as follows,

$$log\left(\frac{p_{ij}(\boldsymbol{X})}{1-p_{ij}(\boldsymbol{X})}\right) = \beta_{0j} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_k X_{kij}$$

$$\beta_{0j} = \beta_0 + u_j$$

Where $p_{ij}(\boldsymbol{X})$ is the probability that census tract 'i' belonging to county 'j' will be classified as unhealthy due to its unique landscape signature metrics and socio-economic predictors. Note that $u_j \sim N(0, \sigma^2)$ is the term that modifies the intercept in the logistic regression model to control for county-specific effects. During the model development process, the statistical significance of u_j is seen as a proxy for the significance of county-level effects on the outcome. Additional predictors can be added at the county level (for example, county land use and SES metrics). However, it is infeasible to add to the current dataset as there are only 21 counties in the study area.

Table 2. Conceptual Framework of multilevel model and associated predictors

	Dependent Variables	Independent Variables
Spatial scale 1 (Census Tract)	{Health outcomes}	Land Use pattern metrics Neighborhood Deprivation Index
Spatial scale 2 (County)		County-level random effects (random intercept, fixed slope model)

Limitations of research design

Four primary limitations in the research design are described below:

1. Limitations of sample size: This research attempts to extract a lot of information through sophisticated modeling techniques. However, the data is fairly limited as it is constrained to 21 counties and a subset of approximately 950 census tracts. Reliability of estimates will have to be carefully examined for their utility in translating to policy recommendations. Furthermore, no causal inferences can be drawn from the study about the mechanisms that mediate the relationship between landscape patterns and outcomes. In the hierarchy of epidemiological studies, this research establishes preliminary associations within reasonable theoretical consistency and expectations in the existing literature. It provides a foundation for further causal studies.
2. Limitations of land use data: The land use dataset used in this study is a hybrid of remotely-sensed data (on-screen photo-interpretation and digitizing of ortho-rectified, high resolution aerial photography) as well supplementary data including ownership information provided by the counties. Classification of remotely-sensed data

introduces errors which might result in misclassification or oversimplification of land use polygons. However, the LandPro dataset is the best available that has consistent land use classification across the entire 21-county region. It is also routinely used by the Atlanta Regional Commission for transportation, environmental and landuse planning. Limitations of cross-sectional study: This research presents a snapshot of data at one point in time. Ideally, a longitudinal or panel data set is better suited to estimate the lag between changes in land use and resulting health impacts.

3. The Modifiable Areal Unit Problem: The research computes landscape patterns at artificially imposed political boundaries (census tracts). While these approximate political jurisdictions where land use planning occurs, they might not represent actual spatial domains of neighborhoods or areas of activity relevant to everyday human experience. It is very plausible that the ideal scale to calculate landscape patterns for health impacts lies in between the county and tract scale. Furthermore, data limitations do not permit the release of health variables at scales smaller than the tract level or those other than political boundaries.

Anticipated research contributions and outcomes

This dissertation creates a consistent, scalable framework to measure land use and its associations with health outcomes. This method is highly replicable and can be used productively across the spectrum of neighborhood, regional and any other theoretically relevant scale. In the short term, it creates a more nuanced form of measurement that lends insight into compositional and configurational characteristics. This approach has largely been unexplored in the built environment and health research. In the long term, it provides a methodological bridge between biocentric and anthropocentric definitions of healthy places. This holds a lot of promise in developing a more holistic approach to sustainable land use planning.

While aspects such as land use mix are known to have associations with health behaviors and outcomes, they are often measured as compositional averages. This research uses several additional metrics that quantify the shapes, distributions and interrelationships to study these associations in greater depth. The projected outcome from this research is the selection of a key set of landscape metrics that provide new insights into the built environment and health relationships. The set of key metrics can provide powerful guidance on future land use planning for healthy places. Statistical modeling is used extract these key variables, providing a quantitatively rigorous framework. This research also demonstrates applications of cutting-edge statistical methods such as multilevel modeling and unsupervised data mining to built environment-health research.

While sprawl metrics measure urban development and land patterns, they provide little insight into tangible measures that can inform physical planning. Landscape metrics offer practical guidance towards sizes, shapes, distributions and juxtapositions towards plan-making. The research generates geometric and other physically significant thresholds significantly missing in the literature. It introduces innovative approaches both theoretically and methodologically. Theoretically, it introduces a new approach to measuring urban form and tests its validity for health research. Methodologically, it provides a comparison between traditional and cutting-edge data mining techniques that have not been well explored in “healthy places” research.

CHAPTER IV

BUILDING THE DATASET

This dissertation analyzes the relationship between three primary sets of variables: 1) Health Outcomes (dependent variable) 2) Landscape Metrics (independent variable) 3) Socioeconomic Index (independent variable). Chapter 4 explains the process of constructing these variables from the raw data.The flow diagram below (Figure 18) describes the process and methodology used to convert the raw data into the desired formats for analysis.

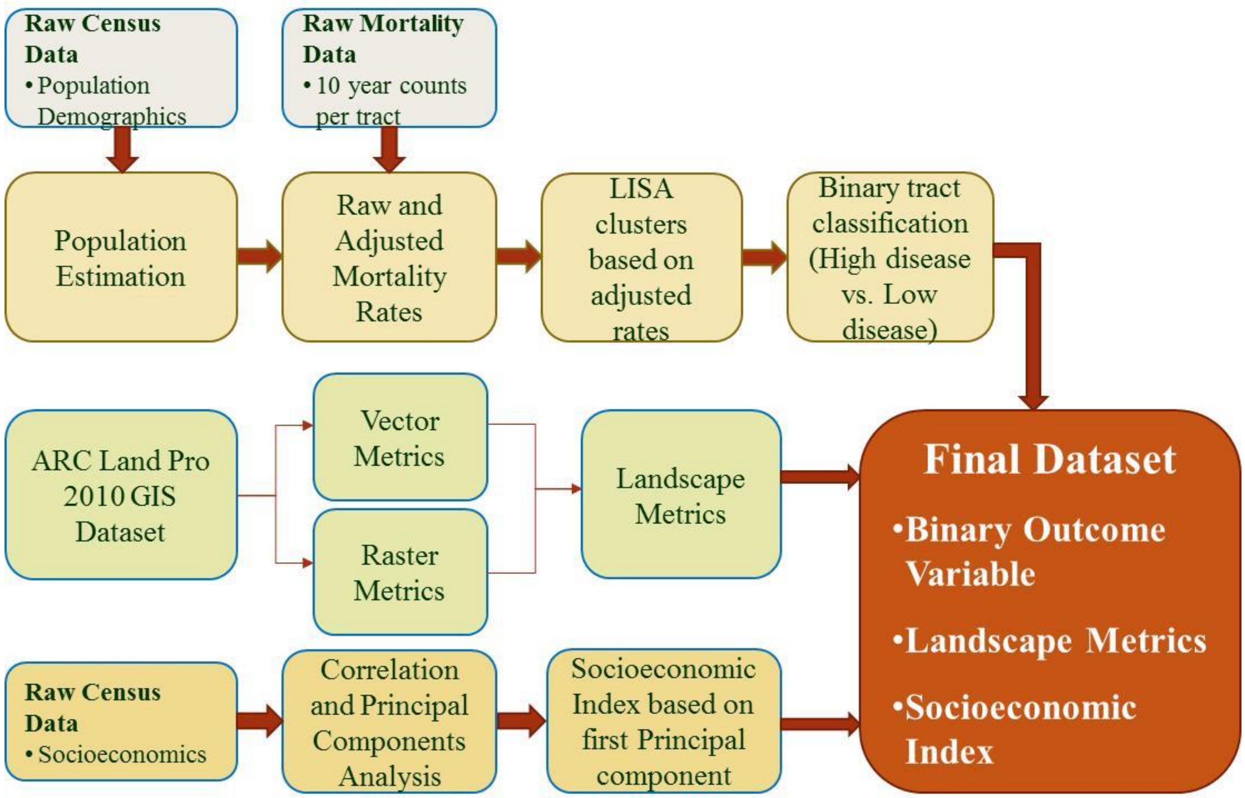


Figure 18. Process flow of building the dataset

Health outcomes: defining and constructing the dependent variable

The primary form of health data used in this research is combined mortality counts for census tracts for the years 2002-2011, stratified across several categories (Table 3).

Table 3. Health Outcomes Data (Source: Georgia Department of Public Health, Office of Health Indicators for Planning (OHIP)

Category	Measures / Indicators	Strata	Years	Spatial units
Mortality	Deaths	• Age • race • sex • cause-specific	Aggregate counts for 2002-2011	County and census tract

The primary aim of this dissertation is to identify landscape and socioeconomic signatures that can be potentially used to classify healthy and unhealthy places. The framework and methodology developed in this dissertation is intended to be generalizable to other problems/datasets that might be relevant for other regional planning issues (mortality rates used in this research is only one measure of health outcomes). This dissertation is thus framed as a classification problem rather than a prediction problem (prediction of mortality rates). A binary outcome variable was best suited for this type of analysis.

Description of the GA DPH dataset

The Georgia Department of Public Health makes health data available through two primary channels—1) The Online Analytical Statistical Information System (OASIS) which is a web-based tool that allows access to publicly available health data and statistics for the state of Georgia, 2) The Public Health Information Portal (PHIP) which is an online data request system permitting requests to data not available on OASIS. The OASIS portal provides access to county-level health outcomes as a continuous variable which can be downloaded as a data table. The mapping tool permits visualization of tract-level health data for counts only. However, this data is provided in the form of numerical ranges (categorical data). A maximum of five data classes are available either under the 'Natural Breaks' or 'Quantiles' classification. The maps can be saved but the data cannot be downloaded.

This dissertation examines the impact of land use patterns on health outcomes at county and tract scales, necessitating the need for tract-level health data available as actual values and not ranges. This enables the calculation of age-adjusted rates, normalizing the health data and making tracts comparable. Specific indicators (health outcomes impacted by environmental conditions) were selected from the warehouse of indicators made available by GaDPH. Data was requested through the PHIP system. After a careful evaluation of the data request, mortality data for selected outcomes were released in the form of aggregate counts for the years from 2002-2011 for all census tracts in the State of Georgia. This also circumvents the issue of having very small counts in certain areas as well as any privacy issues with small-area data.

Data on health outcomes will be analyzed at the county and census tract scales. The 10-year aggregate counts have also been found to be more reliable than annual counts for smaller geographic units such as tracts. When aggregated over several years, these counts provide more realistic estimates of events (smoothed estimates) rather than annual counts which might be an anomaly rather than a trend (Cromley and Lafferty, 2012). Mortality rates and underlying causes of death also represent objective measures as against morbidity or self-reported data which might be inaccurate and biased. More importantly, the mortality counts are available for several obesity and lifestyle related chronic health conditions such as cancer, respiratory illnesses and cardiovascular disease.

The dissertation definitely utilizes the best available data at the best possible scale where multiple datasets could be reconciled. While the definition of neighborhood/local scale can vary considerably across regions, the use of data at standard geographic units such as tract and county, poses several advantages. First, it enables the expansion of the research framework presented here to other regions and makes the results comparable. Second, it permits the possibility of adding other secondary data from publicly available sources (County Health Rankings, BRFSS). Health data at the census tract level is the smallest unit at which health departments are comfortable in publishing data when feasible. This trend is evident in initiatives such as the 500 Cities Project where the CDC is working towards releasing data for 27 chronic disease measures for the 500 largest American cities at the census tract level. Estimates of socioeconomic measures from the American community Survey are also more reliable at the census tract level.

Selecting and grouping mortality categories

The raw data also had numerous categories of disease types. The 10-year mortality counts provided for each tract (951 tracts in total) was further divided into numerous strata (age, gender, race). Data was summarized over all these strata (except age) and total counts for each tract were created (using pivot tables in Excel). The age strata were reconciled with census data so age-adjusted rates could be computed. The next stage was to choose the disease outcomes of interest to be analyzed. The primary criteria used to select these disease types was that they were impacted by environmental conditions. Among these, outcomes with the highest contribution to total mortality in the study area were further selected. Related diseases (based on shared risk factors) were then further grouped into broader categories for analysis. Table 4 below shows the relative contribution of the particular disease outcome to total mortality in the study area (%Total column). The final four disease classifications considered for analysis were:

Heart Disease

Lung Cancer

COPD

Diabetes

Table 4. Summaries of disease counts and their contribution to overall mortality in the Atlanta Region. The table also shows groupings of disease categories.

Cause Of Death	Deaths	%Total	ClusterID
Accidental Discharge of Firearms	80	0.03%	exclude
Accidental Drowning and Submersion	462	0.16%	exclude
Accidental Exposure to Smoke, Fire and Flames	381	0.13%	exclude
Accidental Poisoning and Exposure to Noxious Substances	3734	1.29%	exclude
Acute Rheumatic Fever and Chronic Rheumatic Heart Diseases	246	0.09%	exclude
Alcoholic Liver Disease	966	0.33%	exclude
All COPD Except Asthma	12920	4.47%	**COPD**
All Other Chronic Liver Disease and Cirrhosis	1960	0.68%	exclude
All Other Diseases of the Genitourinary System	1681	0.58%	exclude
All Other Diseases of the Nervous System	3864	1.34%	exclude
All Other Endocrine, Nutritional and Metabolic Diseases	2943	1.02%	**Diabetes**
All Other Mental and Behavioral Disorders	12231	4.23%	**Mental_Disease**
All Other Unintentional Injury	1926	0.67%	exclude
Alzheimers Disease	6991	2.42%	exclude
Anemias	618	0.21%	exclude
Assault (Homicide)	3742	1.29%	exclude
Asthma	463	0.16%	**COPD**
Cerebrovascular Disease	15624	5.40%	**Heart_Disease**
Certain Conditions Originating in the Perinatal Period	2753	0.95%	exclude
Congenital Malformations, Deformations and Chromosomal Abnormalities	1512	0.52%	exclude
Diabetes Mellitus	6499	2.25%	**Diabetes**
Diseases of the Musculoskeletal System and Connective Tissue	1499	0.52%	exclude
Essential (Primary) Hypertension and Hypertensive Renal, and Heart Disease	10011	3.46%	**Heart_Disease**
Falls	2789	0.96%	exclude
Hodgkins Disease	158	0.05%	exclude
Human Immunodeficiency Virus (HIV) Disease	3511	1.21%	exclude
Infections of Kidney	73	0.03%	exclude
Influenza	103	0.04%	exclude
Intentional Self-Harm (Suicide)	4860	1.68%	exclude
Ischemic Heart and Vascular Disease	33880	11.71%	**Heart_Disease**
Legal Intervention	87	0.03%	exclude
Leukemia	2361	0.82%	**Cancer_Other**
Malignant Melanoma of the Skin	880	0.30%	**Cancer_Other**
Malignant Neoplasm of Bladder, Kidney, and Renal Pelvis	2753	0.95%	**Cancer_Other**
Malignant Neoplasm of Pancreas	3650	1.26%	**Cancer_Other**
Malignant Neoplasm of Prostate, and Testis	3143	1.09%	**Cancer_Other**
Malignant Neoplasm of Stomach	1416	0.49%	**Cancer_Other**
Malignant Neoplasm of the Breast	5644	1.95%	**Cancer_Other**
Malignant Neoplasm of the Cervix Uteri, Uterus, and Ovary	3241	1.12%	**Cancer_Other**
Malignant Neoplasms of Colon, Rectum and Anus	6069	2.10%	**Cancer_Other**
Malignant Neoplasms of Lip, Oral Cavity, Pharynx, and Esophagus	2328	0.80%	**Cancer_Other**
Malignant Neoplasms of Liver and Intrahepatic Bile Ducts	1909	0.66%	**Cancer_Other**
Malignant Neoplasms of Meninges, Brain and Other Parts of Central Nervous Sys	1640	0.57%	**Cancer_Other**
Malignant Neoplasms of the Trachea, Bronchus and Lung	17605	6.09%	**Cancer_Lung**
Meningococcal Infection	12	0.00%	exclude
Mental and Behavioral Disorders due to Psychoactive Substance Use	1449	0.50%	exclude
Motor Vehicle Crashes	6059	2.09%	exclude
Nephritis, Nephrotic Syndrome and Nephrosis	6959	2.41%	exclude
Non-Rankable	68177	23.57%	exclude
Parkinsons Disease	1927	0.67%	exclude
Pneumonia	5916	2.04%	**Pneumonia**
Pregnancy, Childbirth and the Puerperium	168	0.06%	exclude
Septicemia	5918	2.05%	exclude
SIDS	596	0.21%	exclude
Suffocation	610	0.21%	exclude
Tuberculosis	69	0.02%	exclude
Unknown	228	0.08%	exclude
Grand Total	**289294**		

Calculating age-adjusted mortality rates

The raw data provided by GaDPH does not account for the underlying age distributions that contribute to the mortality counts within a census tract. For example, a tract with a larger number of older residents could have an excess number of age-related mortality rather than attributing that to some other external environmental risk factor. Age-adjusted mortality rates account for these differential age distributions and make census tracts with different age structures more comparable. Age-adjusted rates were calculated using the following equation (National Cancer Institute):

$$AArate_{x-y} = \sum_{i=x}^{y} \left[\left(\frac{count_i}{pop_i} \right) \times 100{,}000 \times \left(\frac{stdpop_i}{\sum_{j=x}^{y} stdpop_j} \right) \right]$$

$AArate_{x\text{-}y}$ = Age Adjusted rate for the tract between age 'x'and 'y'

Pop_i = Population in age group 'i'

$Count_i$ = Count of deaths in the age group 'i'

$Stdpop_i$ = Standard population within age group "i"

$Stdpop_j$ = Sum of population across all age groups

The process of age-adjusting involves the calculation of crude rates of each age group and then calculating a weighted sum of the crude rates based on a comparative age-distribution of a "standard population" (in this case, the age distribution of the Atlanta MSA region). Census data for 2000, 2010 and population estimates for intermediate years from the American Community Survey (ACS) were used to estimate the population at risk. Interpolation was used to estimate population totals for years where there was no data. In other words, person-years were calculated to match the event counts over the 10 year period. While the dissertation uses well-documented methods to calculate the age-adjusted rates ad all intermediate components with the best available data, some uncertainty is associated with the calculated rates (Anselin et al, 2006; Case Western Reserve University, 2012).

Census tract population data between 2000 and 2010 decennial census are not longitudinally comparable as geographical boundaries and locations of census tracts change over time. Population data for intermediate years are also only estimates available via the American Community Survey. This introduces one of the main sources of uncertainty in calculating the age-adjusted rates. In order to use a comparable population dataset over time, Year 2000 population data for 2010 tract equivalents from the Longitudinal Tract Data Base (Logan et al, 2012) developed at Brown University as part of the US2010 project was used (http://www.s4.brown.edu/us2010/Researcher/Bridging.htm). When back-calculated, a small number of tracts have very small populations for 2000. The box plots for the adjusted rates (Figure 20) also consistently show two tracts that appear to be extreme outliers. Considering these sources of error, the actual age-adjusted rates may skew a linear regression analysis if used as a continuous variable. Hence, the decision was made to convert the rates into a binary, categorical variable. Figures 21-24 show the age-adjusted mortality rates mapped in deciles to visualize the spatial patterns in the study area.

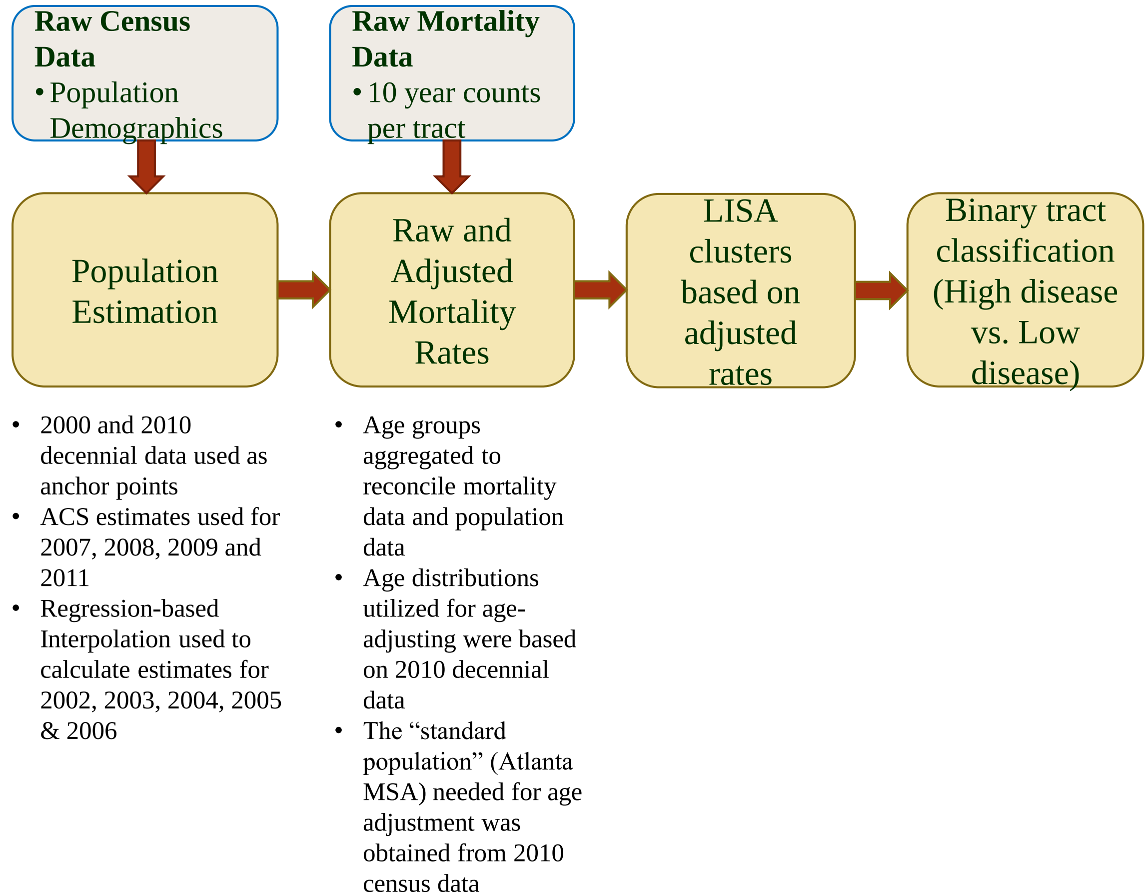


Figure 19. Process used to compute age-adjusted mortality rates

Descriptive statistics and mapping of adjusted rates

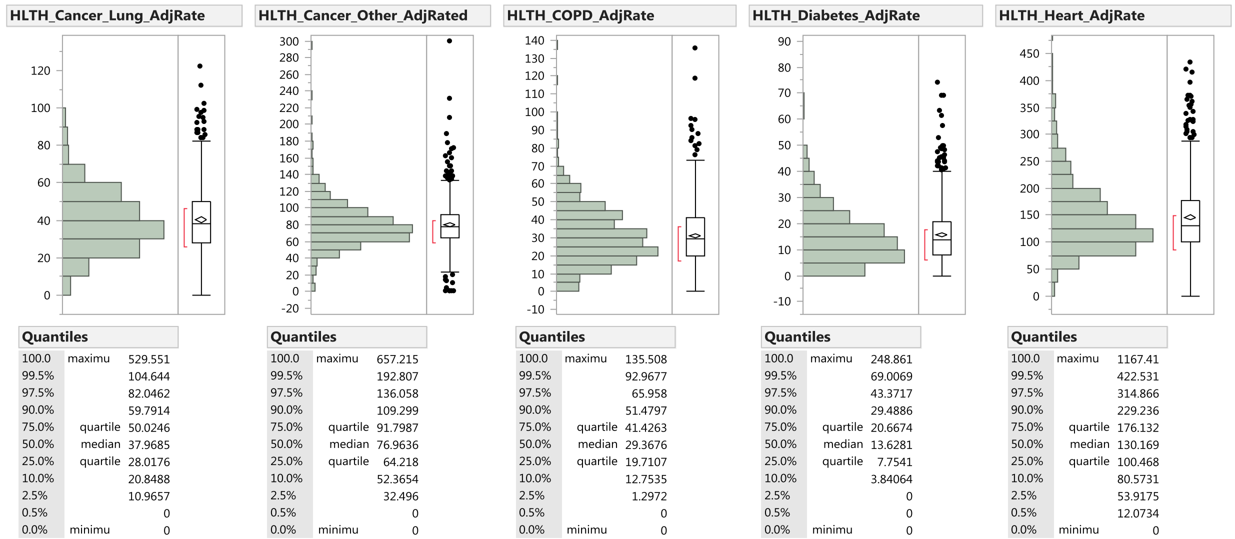
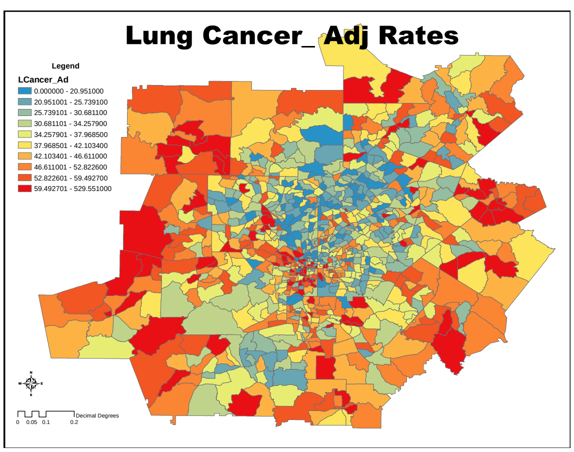


Figure 20. Histograms showing the distributions of age-adjusted death rates of the diseases of interest

<u>Lung Cancer</u>

Map below shows adjusted rates mapped as deciles

Maps on the right show the LISA clusters and their associated significance levels



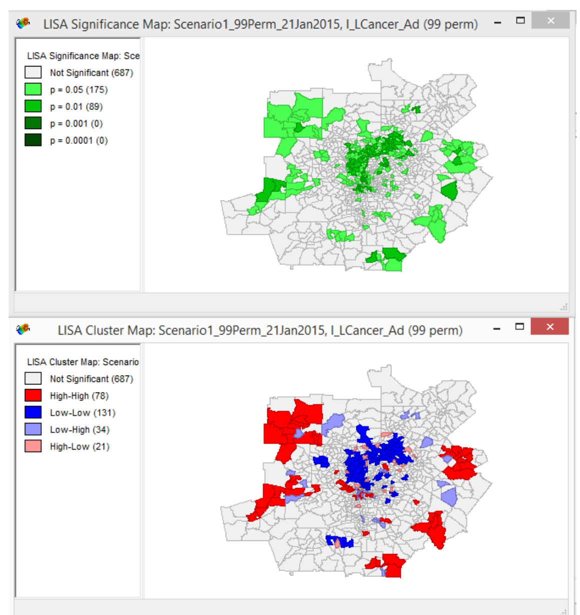
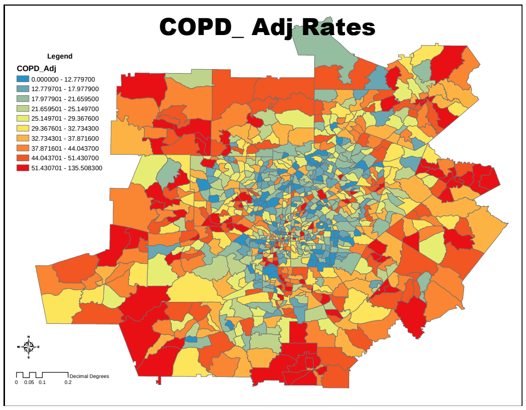


Figure 21. Maps of age-adjusted Lung Cancer mortality rates and associated LISA clusters. The general spatial pattern indicates a ring of higher rates around the periphery of the region.

COPD

Map below shows age adjusted rates mapped as deciles

Maps on the right show the LISA clusters and their associated significance levels



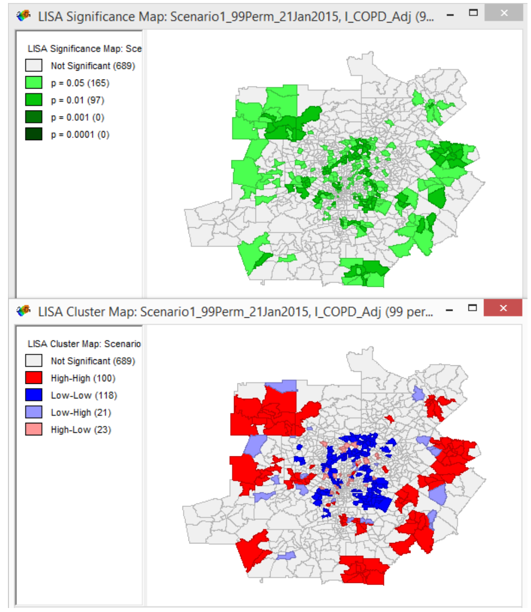
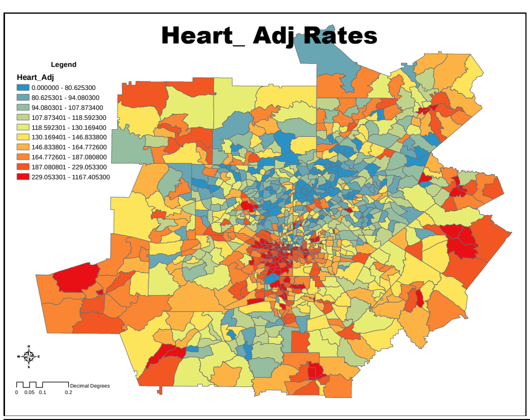


Figure 22. Maps of age-adjusted Lung Cancer mortality rates and associated LISA clusters. The general spatial pattern is similar to Lung Cancer.

Heart Disease

Map below shows age adjusted rates mapped as deciles

Maps on the right show the LISA clusters and their associated significance levels



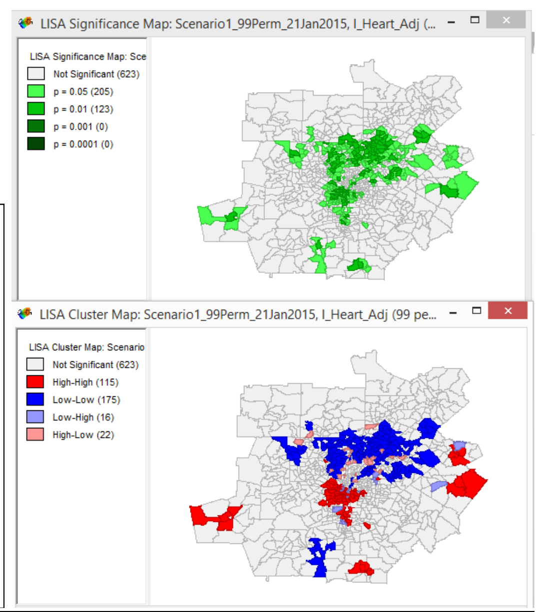
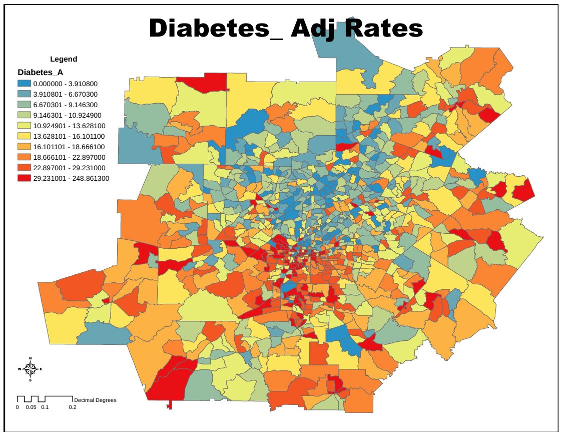


Figure 23. Maps of age-adjusted Heart Disease mortality rates and associated LISA clusters.

Diabetes

Map below shows age adjusted rates mapped as deciles

Maps on the right show the LISA clusters and their associated significance levels



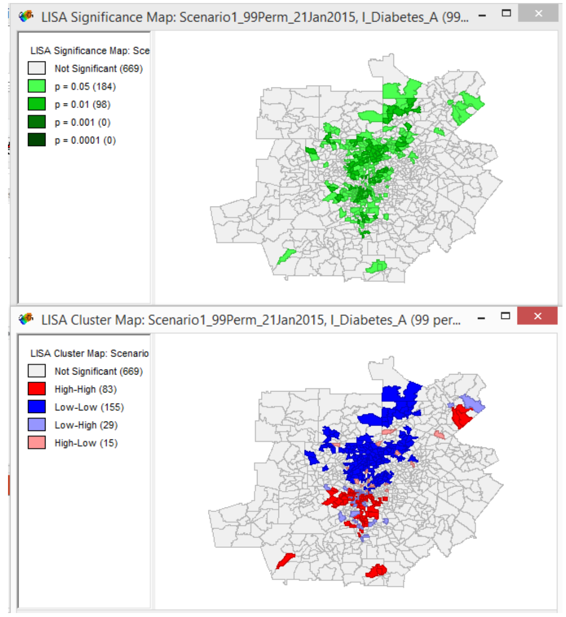


Figure 24.Maps of age-adjusted Diabetes mortality rates and associated LISA clusters.

Generating binary thresholds

The next step was to convert the age-adjusted mortality rates (continuous variable) into a categorical, binary variable. From a framework perspective, it allows the investigation of the important land use and social determinants that contribute to “healthy” or “unhealthy” places, conceptualizing it as a classification problem rather than a predictive modeling problem. In general, age-adjusted rates are sensitive to changes in population demographics and mortality counts and can cause model stability issues if used directly as an outcome variable for linear regression. By treating health outcomes as a classification problem using a carefully-selected threshold, the analysis is more robust and generalizable. This also allows the use of many popular data-mining methods like binary and multilevel logistic regression, which provide model coefficients that are easily interpretable and useful from a policy-making perspective. The mortality rates were initially mapped based on deciles to capture as much fine-grained variation as possible. However, deciding on a threshold that set a cut-off point for low vs high rates would still have been arbitrary. In order to avoid this, a series of steps were executed to develop a more data-driven threshold.

A spatial cluster is conceptually described as an unusual or excessive collection of events which are close together in geographic space after accounting for the heterogeneity in the underlying at-risk population. Alternatively, it can also be defined as a spatial pattern that differs from a random pattern where an event is equally likely to occur at any location (events follow a uniform distribution and are independent of one another) (Moraga and Montes, 2011; Jacquez, 2008; Waller and Gotway, 2004).

Clusters are usually measured through indicators of spatial autocorrelation. Global indexes of spatial autocorrelation provide a summary of the entire study area and in essence answer the question “Is there spatial clustering in the area or not?”. On the other hand, local indicators of spatial autocorrelation (LISA) go beyond merely suggesting clustering to identifying local aggregations of “mutually similar deviations from the overall mean regional count or proportion” (“Where are the spatial clusters?”). LISA clusters (local Moran’s I statistic) for the selected disease rates were generated using GeoDa software (Anselin, Syabri and Kho, 2005). GeoDa generates the following types of clusters:

- Not significant (Areas that are not significant at a significance level of 0.05)
- High-High (High values surrounded by high values)
- Low-Low (Low values surrounded by low values)

- Low-High (Low values surrounded by high values)
- High-Low (High values surrounded by low values)

For the first round of analysis, tracts belonging to the high-high clusters and tracts with the high values in the high-low clusters were combined and coded as unhealthy (1) and the rest were coded as healthy (0). Preliminary binary logistic regression models were run with this classification. However, model diagnostics revealed class imbalance issues as there were too few unhealthy tracts. Also, several tracts that had high values similar to the tracts in the high-high clusters (high mortality rates) were being misclassified as healthy because they were not a part of a cluster.

The presence of clusters is by themselves less important to the models (also indicates that spatial autocorrelation might not be of significant concern in the area). However, what the LISA clusters provide are high rates that deviate significantly from the regional mean. Thus, the lowest rate represented in all the clusters was derived as an objective threshold to separate high and low disease rates. All values (inside or outside a cluster) equal to or above this threshold value were reclassified as unhealthy (1) and all those below the threshold were classified as healthy (0). Figure 25 shows the binary thresholds that were generated for the four disease of interest. Tracts shown in red are coded as 1 and tracts shown in blue are coded as 0.

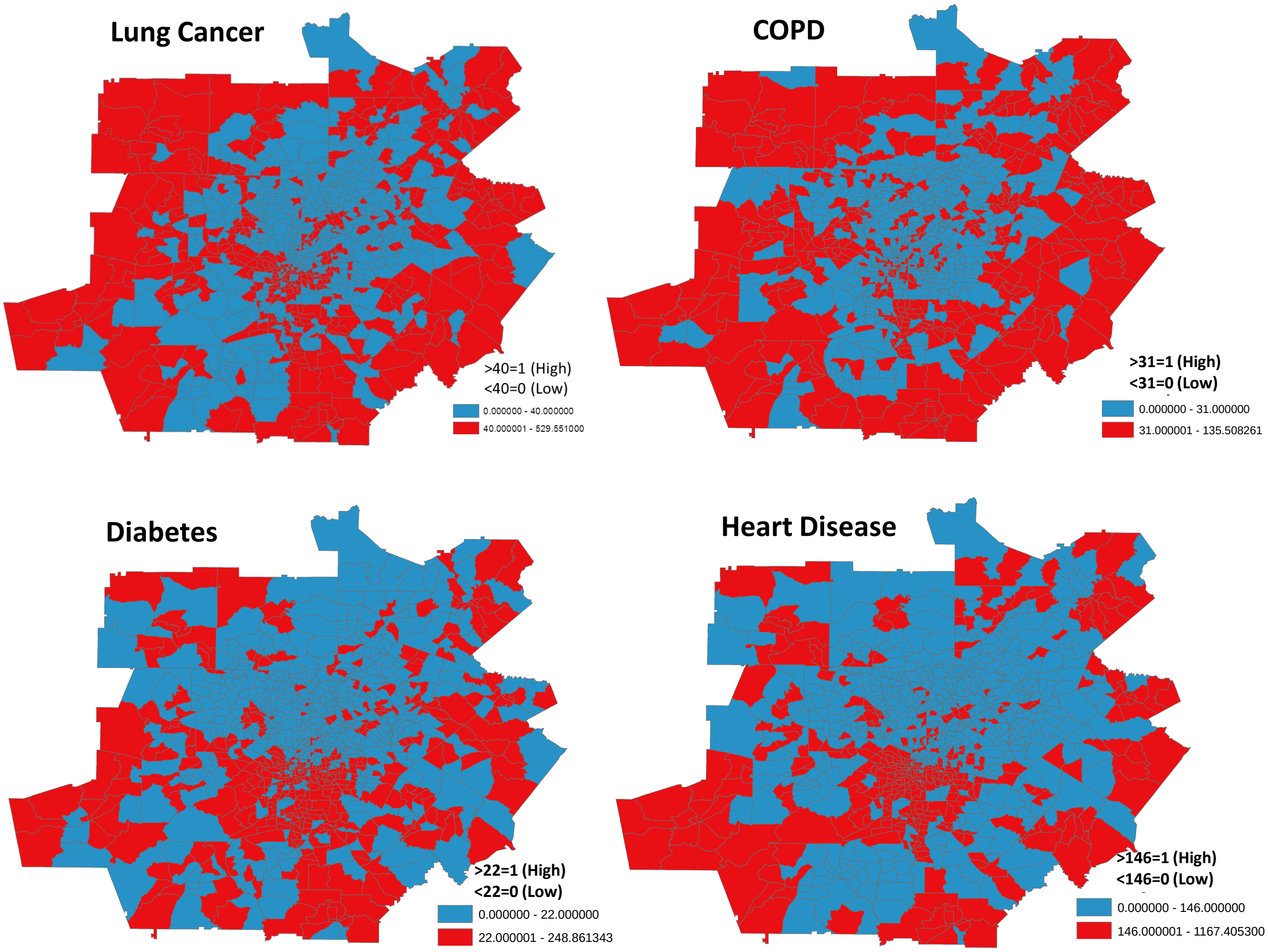


Figure 25. Maps of binary thresholds generated from age-adjusted mortality rates for all four diseases of interest.

Landscape Patterns: defining and constructing the Independent Variables

Class and landscape level metrics were generated using Patch Analyst. Patch analyst is an extension to the ArcGIS® software system that facilitates the spatial analysis of landscape patterns.

Description of the ARC LandPro dataset

The aim of the dissertation is to ascertain the level at which landscape patterns impact health. Landscape patterns are quantified through Landscape Metrics generated from a landuse/landcover dataset. The Atlanta Regional Commission (ARC) publishes the LandPro dataset for the 21 county Atlanta Metropolitan Region that falls under the ARC jurisdiction for regional land use and transportation planning. The 2010 LandPro dataset was used for this dissertation. The LandPro dataset is a hybrid of landcover and land use data.

This GIS dataset combines information from remotely-sensed, orthorectified, aerial imagery and supplemental ownership information. The land cover data is derived from sources such as Aero Surveys of Georgia street atlas, the Georgia Department of Community Affairs (DCA) Community Facilities database and the USGS Digital Raster Graphics (DRGs). The land use component is based primarily on ownership information provided by the counties and the City of Atlanta The landuse/cover classification system is adapted from the USGS (Anderson) classification system, incorporating a mix of level I, II and III classes. There are a total of 25 categories in ARC's landuse/cover system (described below), 2 of which are used only for landuse designations: Park Lands (Code 175) and Extensive Institutional (Code 125). The other 26 categories can describe landuse and/or landcover, and in most cases will be the same. The LU code will differ from the LC code only where the Park Lands (Code 175) and Extensive Institutional (Code 125) land holdings have been identified from collateral sources of land ownership. These lands represent areas most likely to be conserved to left undeveloped and they are controlled by the government (ARC LandPro metadata).

Landscape Metrics

Table 5 shows the categories and measures of Landscape Pattern Metrics that were generated using Patch Analyst in Arc GIS. A more detailed list and description of each metric is shown in Table 6.

Table 5. Landscape Pattern Metrics grouped by the aspect of Landscape Pattern that they measure (also referred to as domain in the document)

Type of Metric (Domain of Landscape Pattern)	Measures
Area Metrics	• Class Area • Total Landscape Area
Shape Metrics	• Mean Shape Index • Area Weighted Mean Shape Index • Mean Perimeter-Area Ratio • Mean Patch Fractal Dimension • Area Weighted Mean Patch Fractal Dimension
Patch Density and Size Metrics	• No. of patches • Mean patch size • Median Patch Size • Patch Size Coefficient of Variation • Patch Size Standard Deviation
Edge Metrics	• Total Edge • Mean Patch Edge • Contrasted Weighted Edge
Diversity and Interspersion Metrics	• Mean Nearest Neighbor Distance • Mean Proximity Index • Interspersion Juxtaposition Index • Shannon's Diversity Index • Shannon's Evenness Index

Table 6. List of Landscape Metrics generated in Patch Analyst and their definitions.

Metric	**Description**
Class Area (CA)	Sum of areas of all patches belonging to a given class.
Landscape Area (TLA)	Sum of areas of all patches in the landscape
Percentage of Landscape (ZLAND)	When analyzing by class, ZLAND is the percentage of the total landscape made up of the corresponding class (patch type).
Number of Patches (NumP)	Total number of patches in the landscape if "Analyze by Landscape" is selected, or Number of Patches for each individual class, if "Analyze by Class" is selected.
Largest Patch Index (LPI)	The LPI is equal to the percent of the total landscape that is made up by the largest patch. When the entire landscape is made up of a single patch, the LPI will equal 100. As the size of the largest patch decreases, the LPI approaches 0.
Mean Patch Size (MPS) Median Patch Size (MedPS)	Average patch size. The middle patch size, or 50th percentile
Patch Size Standard Deviation (PSSD)	Standard Deviation of patch areas.
Patch Size Coefficient of Variance (PSCoV)	Coefficient of variation of patches.
Total Edge (TE)	Perimeter of patches.
Edge Density (ED)	Amount of edge relative to the landscape area
Mean Patch Edge (MPE)	Average amount of edge per patch.
Contrasted Weighted Edge Density (CWED)	CWED is a measure of density of edge in a landscape (meters per hectare) with a user-specified contrast weight. CWED is equal to 0 when there is no edge in the landscape, in other words the whole landscape and it's border are made up of a single patch. It's value increases as the amount of edge in the landscape increases and/or as the user increases the contrast weight.

Table 6 continued	
Landscape Shape Index (LSI)	LSI is the total landscape boundary and all edge within the boundary divided by the square root of the total landscape area (square meters) and adjusted by a constant (circular standard for vector layers, square standard for rasters). The LSI will increase with increasing landscape shape irregularity or increasing amounts of edge within the landscape.
Double Log Fractal Dimension (DLFD)	DLFD is a measure of patch perimeter complexity. It nears 1 when patch shapes are 'simple', such as circles or squares and it approaches 2 as patch shape perimeter complexity increases.
Mean Perimeter-Area Ratio (MPAR)	Shape Complexity.
Mean Shape Index (MSI)	Shape Complexity. MSI is equal to 1 when all patches are circular (for polygons) or square (for rasters (grids)) and it increases with increasing patch shape irregularity. MSI = sum of each patch's perimeter divided by the square root of patch area (in hectares) for each class (when analyzing by class) or all patches (when analyzing by landscape), and adjusted for circular standard (for polygons), or square standard (for rasters (grids)), divided by the number of patches.
Area Weighted Mean Shape Index (AWMSI)	AWMSI is equal to 1 when all patches are circular (for polygons) or square (for rasters (grids)) and it increases with increasing patch shape irregularity. AWMSI equals the sum of each patch's perimeter, divided by the square root of patch area (in hectares) for each class (when analyzing by class) or for all patches (when analyzing by landscape), and adjusted for circular standard (for polygons), or square standard (for rasters (grids)), divided by the number of patches. It differs from the MSI in that it's weighted by patch area so larger patches will weigh more than smaller ones.
Mean Patch Fractal Dimension (MPFD)	Shape Complexity. Mean patch fractal dimension (MPFD) is another measure of shape complexity. Mean fractal dimension approaches one for shapes with simple perimeters and approaches two when shapes are more complex.

Table 6 continued	
Area Weighted Mean Patch Fractal Dimension (AWMPFD)	Shape Complexity adjusted for shape size. Area weighted mean patch fractal dimension is the same as mean patch fractal dimension with the addition of individual patch area weighting applied to each patch. Because larger patches tend to be more complex than smaller patches, this has the effect of determining patch complexity independent of its size. The unit of measure is the same as mean patch fractal dimension.
Mean Nearest Neighbor (MNN)	Measure of patch isolation. The nearest neighbor distance of an individual patch is the shortest distance to a similar patch (edge to edge). The mean nearest neighbor distance is the average of these distances (meters) for individual classes at the class level and the mean of the class nearest neighbor distances at the landscape level.
Interspersion Juxtaposition Index (IJI)	Measure of patch adjacency. Approaches zero when the distribution of unique patch adjacencies becomes uneven and 100 when all patch types are equally adjacent.
Mean Proximity Index (MPI)	Measure of the degree of isolation and fragmentation. Mean proximity index is a measure of the degree of isolation and fragmentation of a patch. MPI uses the nearest neighbor statistic. The distance threshold default is 1,000,000. If MPI is required at specific distances, select Set MPI Threshold from the main Patch pull-down menu and enter a threshold distance.
Shannon's Diversity Index (SDI)	Measure of relative patch diversity. Shannon's diversity index is only available at the landscape level and is a relative measure of patch diversity. The index will equal zero when there is only one patch in the landscape and increases as the number of patch types or proportional distribution of patch types increases.
Shannon's Evenness Index (SEI)	Measure of patch distribution and abundance. Shannon's evenness index is equal to zero when the observed patch distribution is low and approaches one when the distribution of patch types becomes more even. Shannon's evenness index is only available at the landscape level.

LandPro is a vector dataset with each landuse/landcover class represented as unique polygons. The dissertation required the creation of landscape metrics at two scales, namely, county and census tracts. Landscape Ecologists call these boundaries "landscape boundaries". 2010 tract and county boundaries were intersected with the LandPro dataset in ArcGIS to create two sets of shapefiles for analysis. One set of shapefiles contains 21 counties with each county (landscape level 2) populated with landuse/landcover polygons. The second set of shapefiles contains 951 census tracts (landscape level 1), each populated with landuse/landcover polygons. The discrete polygons representing a landuse/landcover type are called "patches". Patches belonging to the same land use are collectively called a class.

Landscape Metrics were generated using Geographic Information Systems software (ArcGIS and Patch Analyst). Patch Analyst (developed by the Spatial Ecology Program, The Centre for Northern Forest Ecosystem Research, Ontario, Canada) is a freely downloadable add-on to the ArcGIS environment. It is one among the most popular software (others include FRAGSTATS developed at the University of Amherst, MA) used to generate landscape metrics with an added advantage of being able to analyze vector datasets.

Metrics were generated for each landscape scale (county and census tract) both for individual classes (land uses) as well as for the landscape as a whole. While class level metrics measure composition and configuration of individual patches of the same class, landscape level metrics measure composition and configuration for the entire landscape. Table 6 provides a listing of all metrics generated and their landscape levels.

Selecting Land Uses and their Metrics for analysis

The numerous land use types resulted in a voluminous number of landscape metrics. From a theoretical standpoint, land uses and their corresponding metrics were narrowed down for the final analysis. The criteria for selecting the subset of land uses and their metrics were twofold. Land use types known to influence health outcomes were selected based on the literature review. The percentage of landscape area per tract covered by each land use type was also examined. Land uses that consistently contributed insignificant amounts to each tract and the study area as a whole were identified. Figure 26 shows box plots of the metric "% land use area per tract". Land uses that do not occupy significant land area are flagged using the brown stars. The land uses included in the final analysis are listed and described in table 7.

Table 7. Land Uses included in the final analysis and their definitions as listed in the LandPro metadata

Land Use	**Definition**
Agriculture	Agricultural land regularly used to grow field crops, pasture animals or for livestock production.
Commercial	Areas used predominantly for the sale of products and services, including urban central business districts, shopping centers in suburban and outlying areas, commercial strip developments, junk yards and resorts. Secondary structures and supporting areas are all included: office buildings, warehouses, driveways, sheds, parking lots, landscaped areas, waste disposal areas, etc. Commercial areas may include some non-commercial uses too small to be separated out.
Forest	All forested areas of coniferous and/or deciduous trees.
Industrial / Commercial	Industrial and commercial areas that typically occur together or in close functional proximity with one another.

Table 7 continued	
Industrial	Land associated with light or heavy manufacturing.
Institutional Extensive	Public or private land holdings devoted to educational, religious, health, correctional, or military landuse. The Extensive Institutional landuse areas identify the full extent Institutional tracts which are both built-up and non-built-up, and whose undeveloped area is at least 25 acres in size.
Institutional Intensive	The built-up portions of institutional land holdings, including all buildings, grounds and parking lots that compose educational, religious, health, correctional and military facilities.
Limited Access	This category identifies all highways, or portions of highways that are considered "limited access," their approximate right-of-ways, ramps and interchanges.
Park Lands	Local, state, or federal land holdings devoted to preservation, conservation or recreation, as identified from secondary sources.
Parks	Active recreation areas identified from aerial photography, including baseball and other sports fields, tennis courts, swimming pools, camp grounds, parking lots, structures, drives, and trails
Low Density Residential	Areas that have generally been developed for single family residential use, usually with a significant mix of forested or agricultural landcover. These areas often occur on the periphery of urban expansion and are generally characterized by houses on 2 to 5 acre lots.
Reservoirs	Man-made impoundments, often referred to as "lakes" or "ponds," which are persistently covered with water.
High Density Residential	Areas that have predominantly been developed for concentrated single family residential use. These areas occur almost exclusively in urban neighborhoods with houses on lots smaller than 1/4 acre, but may also include mixed residential areas with duplexes and small apartment buildings.
Medium Density Residential	Areas that have predominantly been developed for single family residential use, with or without a significant mix of forested or agricultural landcover. These areas usually occur in urban or suburban zones and are generally characterized by houses on 1/4 to 2 acre lots. This category accounts for the majority of residential landuse in the Region and includes a wide variety of neighborhood types.
Multi-family Residential	Residential areas comprised predominantly of apartment, condominium and townhouse complexes where net density generally exceeds eight units per acre.

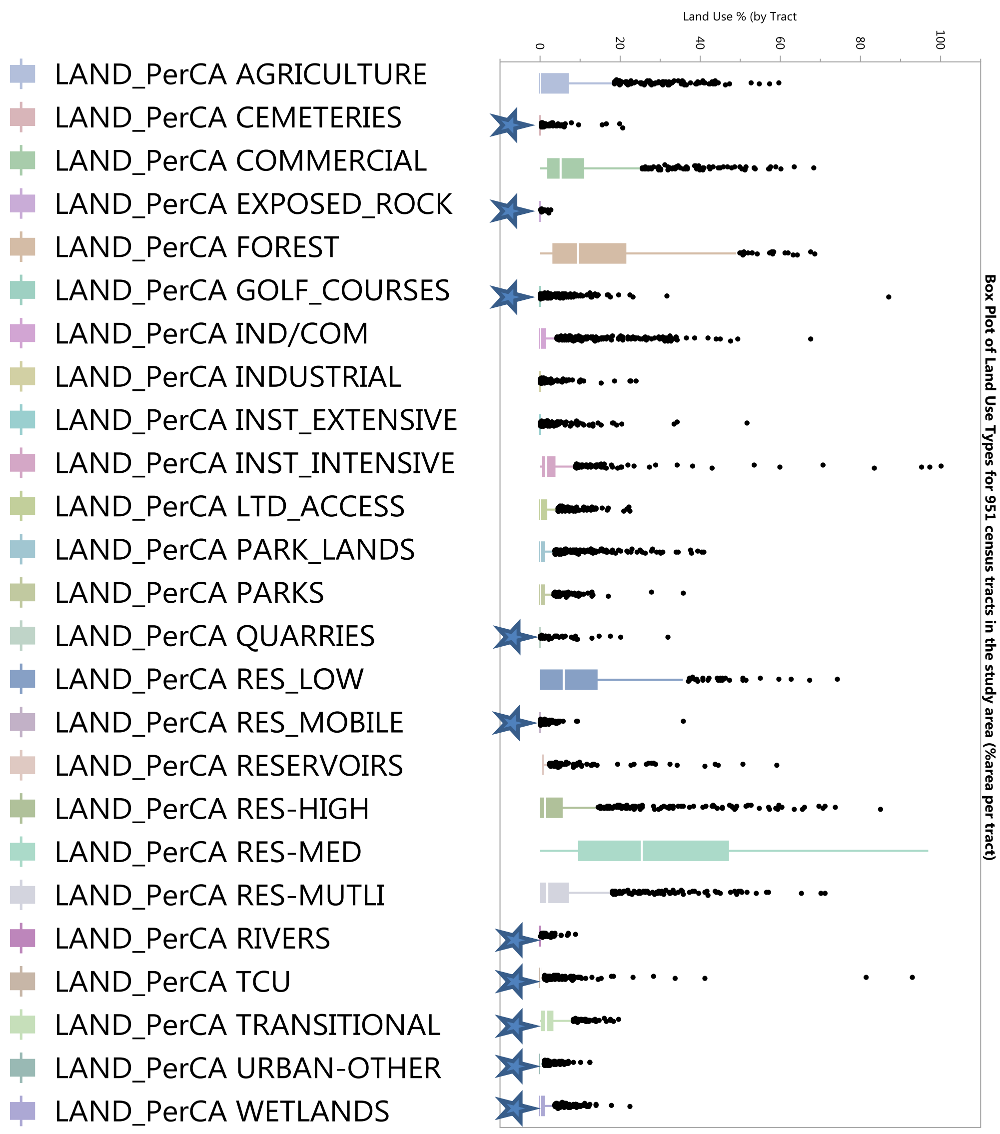


Figure 26. Comparative box plots of land uses showing the percentage of total landscape area they occupy in each tract.

Correlation Analysis

Software packages such as FRAGSTATS and Patch Analyst generate a multitude of landscape metrics that measure various aspects of landscape composition and configuration. However, the literature in Landscape Ecology documents analytical challenges that arise from the abundance of these metrics. Correlation and confounding between metrics makes it challenging to select a parsimonious set of metrics that captures the essence of the composition and configuration of the landscape that is being measured. There are several potential sources of correlation and redundancy. Metrics often measure multiple aspects of landscape structure and composition, making it difficult to isolate the effects of individual metrics (even confounding composition and configuration). Metrics may also be redundant because they measure similar aspects of landscape structure and composition even though they differ mathematically (for example, mean patch size and median patch size are both measures of central tendency).

Principal Components Analysis is a common method employed to reduce the data into a smaller set of independent components. Principal Component Analysis is a common statistical technique used to reduce highly correlated, multi-dimensional data into a smaller set of explanatory variables. These smaller set of variables (principal components) are linear combinations of the original variables. These derived components are completely uncorrelated with each other and can be used as inputs into subsequent regression models (Jolliffe, 2014).

While several studies have attempted to reduce the numerous landscape metrics into a parsimonious set, there is lack of consistency among the number, composition and types of components identified. This potentially implies that landscape structure is not universally describable; rather it is place-dependent.

An initial analysis of the landscape metrics revealed similar correlations. Figure 27 illustrates correlations between landscape metrics for Agriculture and Industrial land uses as examples.

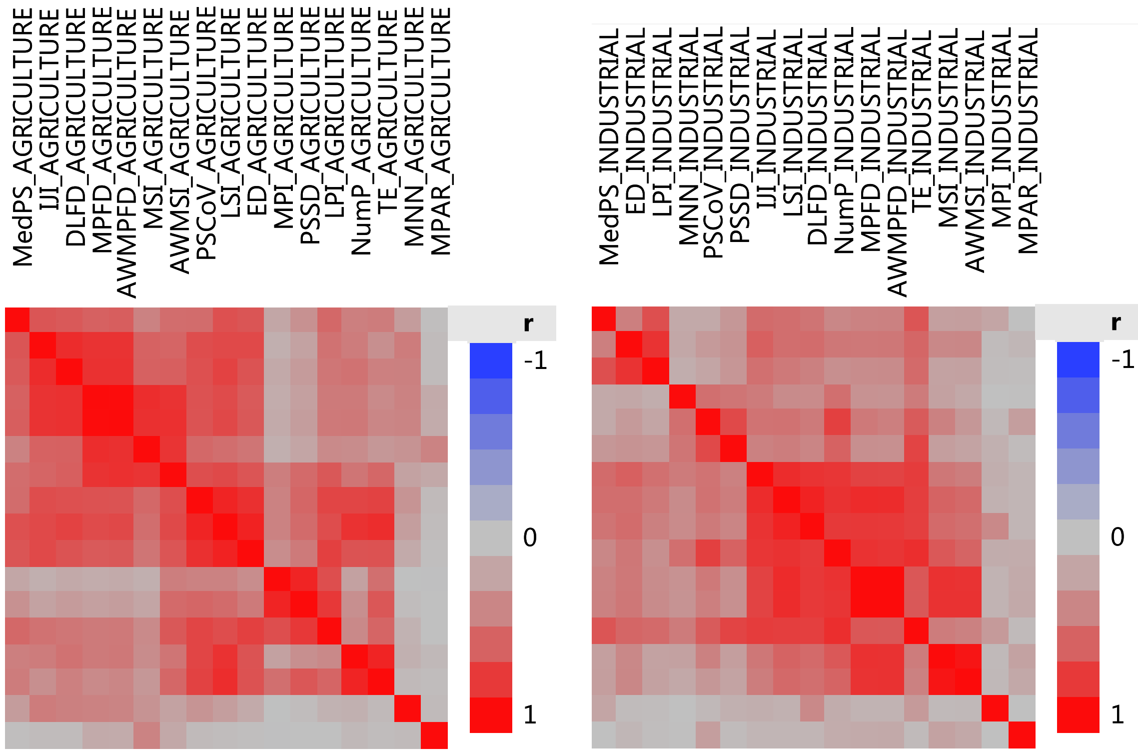


Figure 27. Correlation Matrix of Landscape Metrics

Generating the matrix of Land Use metrics and Indices

Both Principal Components Analysis (PCA) and Factor Analysis (FA) methods were explored for data reduction purposes. The results of the FA was inconsistent in that metrics did not lend themselves to consistent groupings across land use types. Therefore, the decision was made to use PCA to create land use indices to be used in the analytical models.

The purpose of the land use indices was twofold. The first purpose was to reduce the total number of variables to be used in the analysis (also simultaneously eliminate the multicollinearity issue) and the second was to eliminate land uses that were not significantly correlated with health outcomes. Two sets of land use indices were created for modeling purposes. The first land use index (index level 1) was a composite index of all the landscape metrics combined. Due to the high correlation between metrics, the eigen vector of the first principal component tends to explain most of the variation. Thus the level 1 index for each of the land uses was comprised of the first principal component.

The second set of land use indices were created after the first round of analysis. Land uses that emerged as significant were identified and level 2 indices were constructed. The level 2 set are disaggregated indices for each of the significant land uses from level 1. For each of the significant land uses, three level 2 indices were created based on domains identified in the Landscape ecology literature. Each domain represents a different aspect of landscape structure. These domains were *Landscape Geometry (GEO), Landscape Shape (SHA), and Landscape Interspersion (INT) (domains based on FRAGSTATS manual).* While the land use types change based on the disease outcome being analyzed, the domains for the disaggregated index stay the same. Figure 28 below shows the final matrix of land uses and indices constructed for further analysis. GEO contains variables that measure size, perimeter, count and distribution (configuration) of patches for each land use type. SHA contains variables that measure shape complexity of patches. INT is an explicit measure of configuration and measures spatial distribution and adjacencies of patches belonging to the different land use types.

					Types of Land Use														
		Feature ID	Feature Description	Feature Type	AGRICULTURE	COMMERCIAL	FOREST	IND/COM	INDUSTRIAL	INST_EXTENSIVE	INST_INTENSIVE	LTD_ACCESS	PARKS	RES_LOW	RES_MOBILE	RESERVOIRS	RES-HIGH	RES-MED	RES-MUTLI
Used in composite Landscape Index (Index)	Geometry Index (GEO)	**MedPS**	Median patch Size	Geometry															
		PSSD	Patch Size Standard Deviation	Geometry															
		PSCoV	Patch Size Coefficient Of variation	Geometry															
		NumP	Number of Patches	Geometry															
		LPI	Largest Patch Index	Geometry															
		TE	Total Edge	Geometry															
		ED	Edge Density	Geometry															
	Shape Index (SHA)	**MPAR**	Mean Perimeter Area Ratio	Shape															
		LSI	Landscape Shape Index	Shape															
		MSI	Mean Shape Index	Shape															
		MPFD	Mean Patch Fractal Dimension	Shape															
		AWMSI	Area Weighted Mean Shape Index	Shape															
		AWMPFD	Area Weighted Mean Patch Fractal Dim.	Shape															
		DLFD	Double Log Fractal Dimension	Shape															
	INT	**MPI**	Mean Proximity Index	Interspersion															
		IJI	Interspersion & Juxtaposition Index	Interspersion															
		MNN	Mean Nearest Neighbour	Interspersion															
		SDI	Shannon Diversity Index	Diversity	Landscape Level, one value per tract														
		SEI	Shannon Evenness Index	Diversity	Landscape Level, one value per tract														
		SOCIAL	Social Index at Tract Level	Social	Calculated from social indices, one value per tract														

Figure 28. Matrix of final variable set to be used for modeling

Socioeconomics: defining and constructing the Mediating Variables

Both individual and neighborhood socioeconomic status (SES) or socioeconomic deprivation are well-established determinants of health behaviors and outcomes. Research has also shown that area-based deprivation has independent effects on an individuals' health irrespective of their individual socioeconomic status (Solet and Joshi, 1994; link and Phelan, 1996; Berkman and Macintyre, 1997; Krieger et al, 1997; Stafford and Marmot, 2003; Messer et al, 2006). Socioeconomic factors have also been proven as key determinants of excessive mortality, with unhealthy behaviors explaining this association to a significant extent (Singh, G. K. (2003; Sabanayagam and Shankar,2012; Signorello et al, 2014; Mehta et al 2015).

Socioeconomic characteristics such as income, race/ethnicity, education and location of residence have also been shown to be strongly correlated with disparities in healthcare access—another key mechanism along the path from incidence to mortality (Gautam et al, 2014; Archibald and Rankin, 2013). Collectively characterized as neighborhood socioeconomic disadvantage, Kirby and Kaneda (2005) state that socioeconomic deprivation creates physical, service and social impedances to obtaining healthcare. Suggested mechanisms include lack of health insurance, unaffordable healthcare costs, cultural/linguistic factors and beliefs, lack of transportation/vehicle ownership, weak social networks, and inadequate presence of healthcare facilities (Paez et al, 2010; Kirby and Kaneda, 2005).Singh (2003) states that "Community socioeconomic measures describe important aspects of social organization, structure, stratification, or environment, such as socioeconomic deprivation, economic inequality, resource availability, and opportunity structure." While measures such as income, educational attainment or education can be used separately to characterize communities, composite indices that combine these measures capture the multidimensional nature of the SES concept. SES indices are routinely used in the public health literature, particularly for ecological studies that attempt to explain gradients in health outcomes (mortality and morbidity) as a function of SES gradients (Testi et al, 2004; Messer et al, 2006). Measures of derivation also tend to be highly correlated. Deprivation indices are calculated for a pre-defined geographical level, implying that there are independent pathways that separately impact individual health. Area-based measures are also more relevant to this dissertation as area-based deprivation is linked to a lack of resources (material deprivation) which indirectly manifest in the built environment as well. Area-based deprivation measures are prone to the “ecological fallacy” that all individuals living in deprived areas are deprived themselves. However, this dissertation computes a deprivation score at the census tract level which the U.S.Census Bureau defines as “small, relatively permanent statistical subdivision of a county…..Designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time of establishment, census tracts average about 4,000 inhabitants.”

Selecting the socioeconomic variables

Some well-known deprivation indices include the Townsend index of deprivation and the Carstairs deprivation Index (Testi et al, 2004; Messer et al, 2006). However, these indices are created in the context of the United Kingdom and measure social class as an alternative representation of deprivation. They are hard to reproduce in the US as the variables used do not have a direct equivalent in the census and are also calculated at a different scale. However, these indices are based on subtly varying measures of employment, household crowding/tenure, and vehicle ownership. Several deprivation indices have been proposed in the United States context and validated against several health outcomes (Krieger et al, 1997; Diez Roux, 2001; Messer et al, 2006). These tend to be more comprehensive where the seven domains of Poverty/income, racial/ethnic composition, education, employment, and occupation are more consistently represented. housing/crowding, residential stability, economic inequality and racial residential segregation were less commonly utilized (Messer et al, 2006). Messer et al (2006) propose the development of a neighborhood deprivation index using U.S. census tract-level data and principal components analysis (similar framework proposed by Butler, 2013).

This dissertation uses a framework adapted from Messer et al (2006) and Hale et al (2015). Both these frameworks are comprehensive and reproducible indices utilizing census tract data. The Neighborhood Deprivation Index (Messer et al, 2006; Hale et al, 2006) is a multi-dimensional construct representing the domains of poverty, housing, employment, education, car ownership, federal assistance and home occupancy. Both these frameworks select initial variables across the theoretical domains listed above. The final variables (a subset of the initial list) were selected based on a Principal Components Analysis and the final index was a weighted sum of standardized scores. A similar approach will be utilized for this research where an index will be constructed using the following variables shown in table 8. The various income levels included in the index are representative of poverty threshold levels as defined by the census based on household size.

Table 8. Socioeconomic variables used to calculate the Neighborhood Deprivation Index (Data source: Census 2010)

Tract level measure	**County level measure**
Index constructed from the following indicators: • Percent of individuals with income below poverty level • Percent of households with income less than $30,000 • Percent household income LT $10,000 • Percent household income $10000 to $14999 • Percent household income $15000 to $24,999 • Percent household income $25000 to $34,999 • Percent housing costs greater than 35% of household income • Percent of families with female-headed household with dependent children • Percent households with public assistance income • Percent households with no vehicle • Percent unemployed • Percent with less than high school education • Percent vacant housing units • Percent with cash public assistance income • Percent with food stamps/SNAP benefits • Percent using public transportation	GINI Index used in the county-level Hierarchical Clustering (Data source: 2010 County Health Rankings)

Descriptive statistics

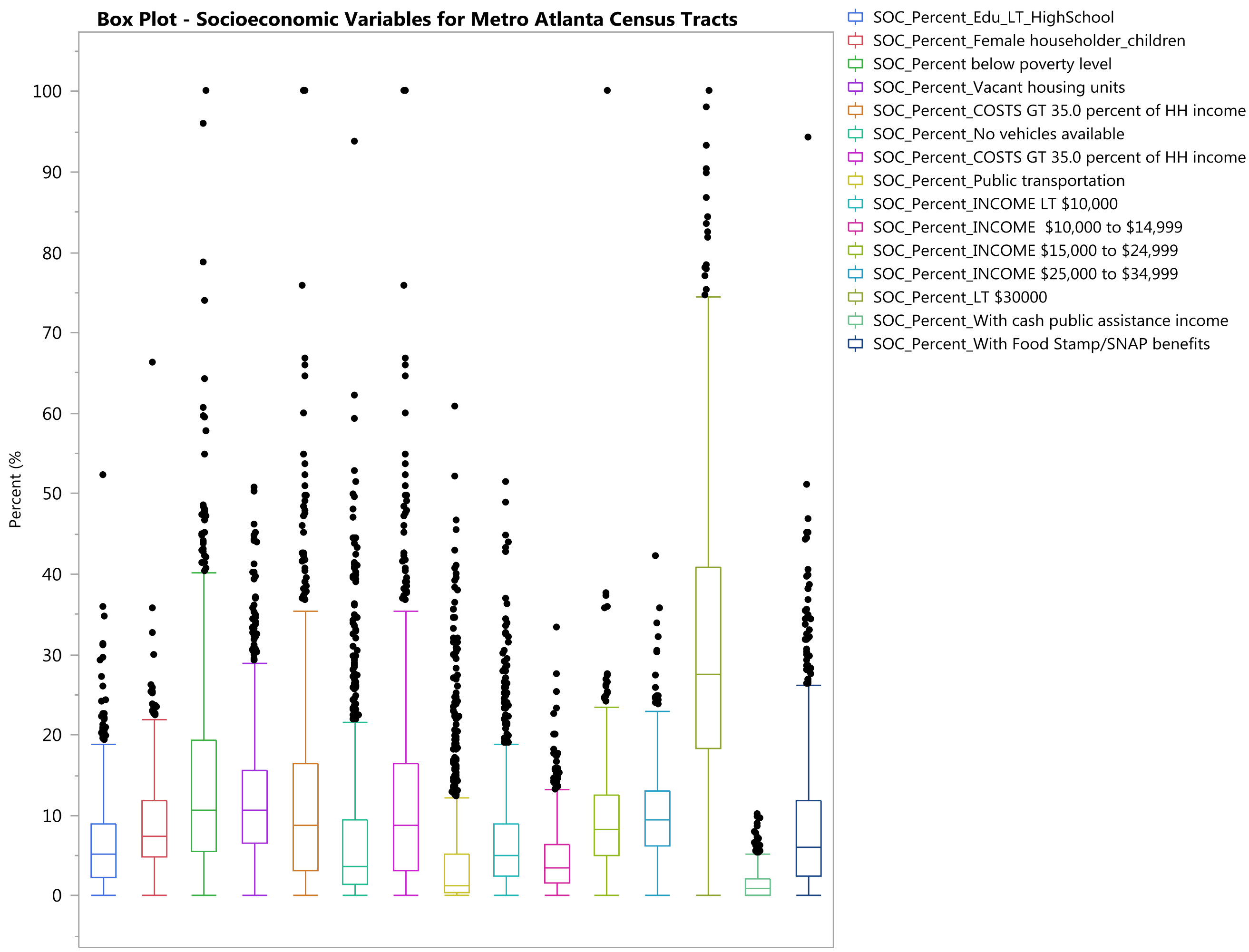


Figure 29. Box plots showing the distributions of socioeconomic variables

Correlation Analysis

The clustering patterns indicate a high degree of correlation between variables as shown by the red grid cells (Figure 30). Principal components Analysis was implemented to aid data reduction and create a single index.

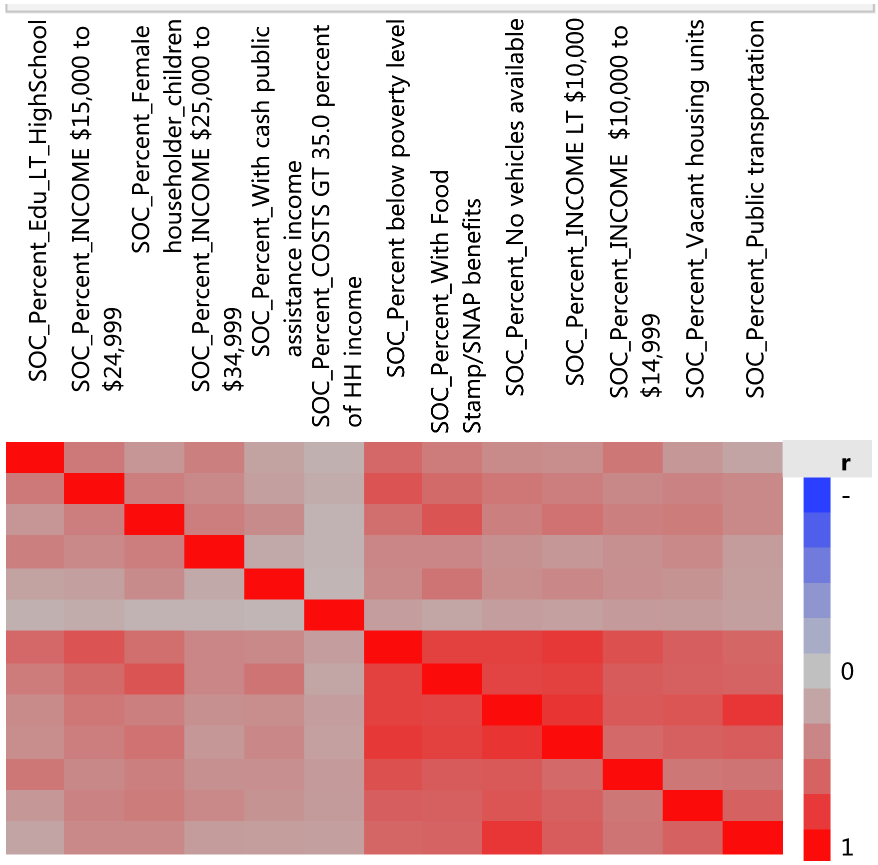


Figure 30. Clustering of correlations among socioeconomic deprivation indicators.

Creating the Neighborhood Deprivation Index using Principal Components Analysis

The first Principal Component explains a majority of the variance and all the variables load on this component. The dominance of the first eigenvalue and a review of similar indices in the literature informed the decision to use the first component to construct the Neighborhood Deprivation Index (Appendix A). Mapping of index values (Figure 31) reflects heuristic knowledge of the socioeconomic patterns that exist in the Atlanta region.

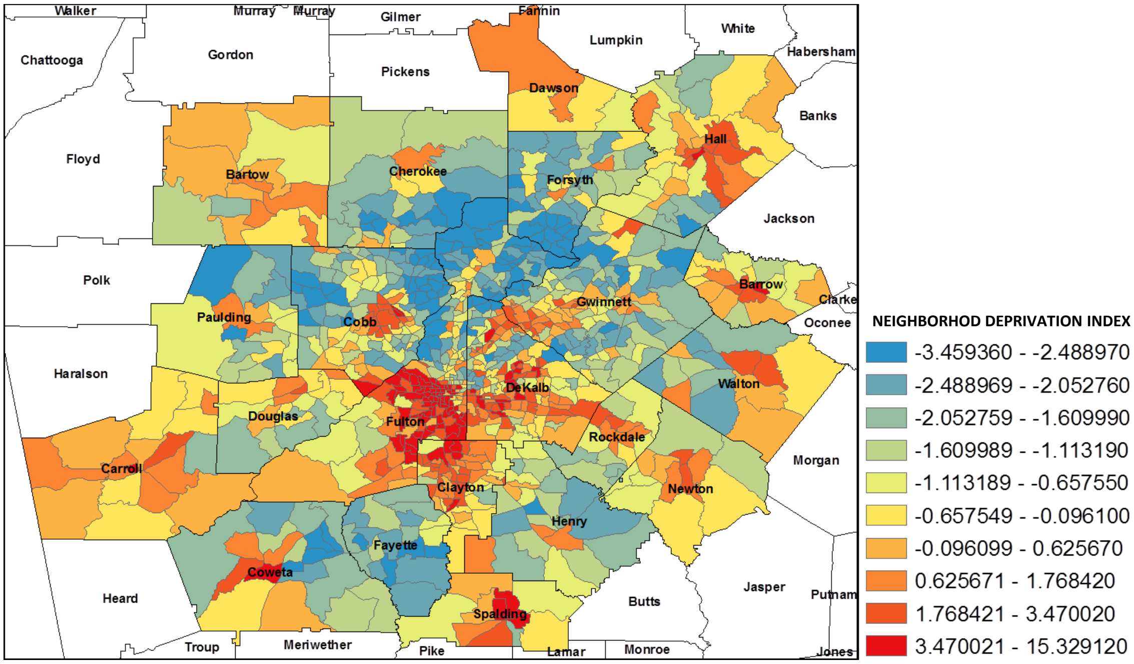


Figure 31. Mapping of the Neighborhood Deprivation Index across tracts in the study area. Low values (blue) indicate lower levels of deprivation and high values (red) conversely indicate higher levels of deprivation.

Tracts with high levels of deprivation are found in the south-central part of the study area. A peripheral ring of tracts has medium-high values for deprivation. Tracts with low deprivation values are generally located in the Northern part of the study area.

Summary

Chapter IV provided methodological details with regard to the construction of the final analytical dataset. This dataset is comprised of the dependent, independent and mediating variables (Figure26 shows the final data matrix). Chapter V provides the analytical process and results from the models utilizing these variables for the outcomes of Lung Cancer and COPD mortality.

CHAPTER V

LUNG CANCER AND LANDSCAPE PATTERNS

Lung Cancer Epidemiology

Nationally, lung cancer continues to be one of the leading types of cancer, both for incidence and mortality (Siegel et al, 2015). The American Cancer Society estimates that new lung cancer cancer cases will account for approximately 13% of all cancer diagnoses in 2015 (Cancer Factsand Figures, 2015). Lung cancer also accounts for the largest number of cancer related and tobacco related deaths in men and women, making it a sugnificant public health concern. The lung cancer five-year survival rate (17.8 percent) is lower than many other leading cancer sites, such as the colon (65.4 percent), breast (90.5 percent) and prostate (99.6 percent). Only 15 percent of cancers are diagnosed at an early stage. While the five-year survival rate for localized lung cncer is 54 percent while the the five-year survival rate for metastatic lung cancer drops steeply to 4 percent.

In the Atlanta metro area, deaths due to lung cancer contributed to approximately 6% of all deaths during the study period between 2002-2011 (second highest cause of death). Figures 32 and 33 illustrate the geographical pattern of age-adjusted Lung Cancer mortality rates in the study region and the subsequent conversion into the binomial dependent variable. The pattern indicates a ring of high mrotality rates along the peripheral tracts of the study area.

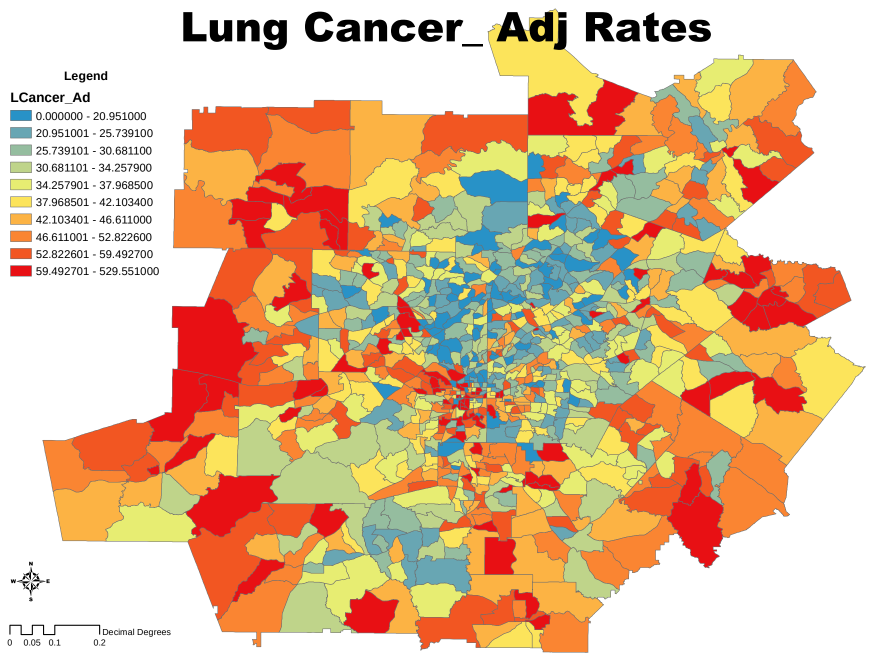


Figure 32. Age adjusted rates for Lung Cancer in the Atlanta Metro area, 2002-2011

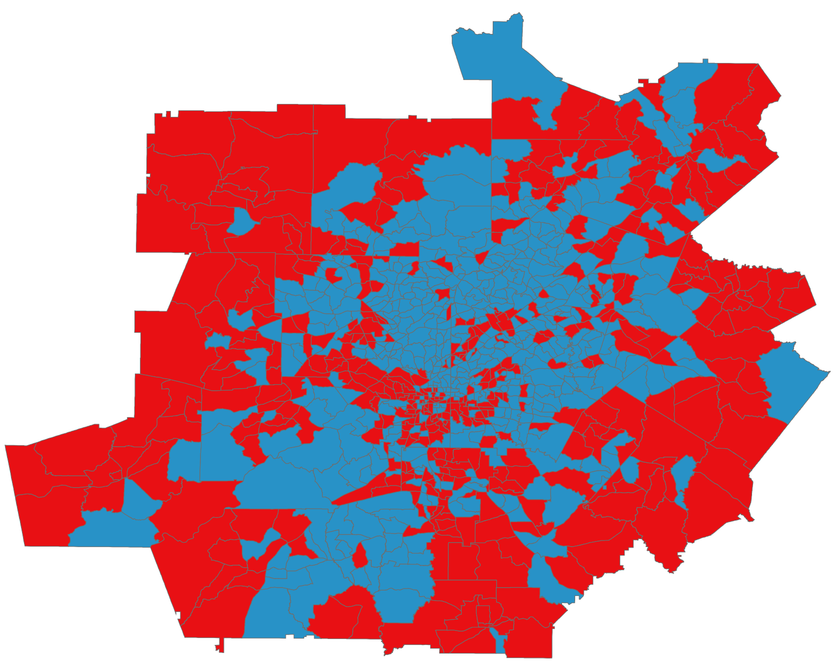


Figure 33. Age-adjusted Lung Cancer mortality rates converted to the binary, categorical dependent variable.

Risk factors for lung cancer

Tobacco smoking continues to be the most important risk factor associated with developing lung cancer. However, several other environmental, occupational and social factors play an important role in cancer incidence and mortality. Household radon gas and scondhand smoke are known siginificant environmental agents. From an occupational standpoint, working in certain industries such as building construction and rubber manufacturing increase the likelihood of being exposed to carcinogens like asbestos, certain metals, organic chemicals, radiation and air pollution (Alberg and Samet, 2003; Cancer Factsand Figures, 2015)

Persistent disparities in cancer incidence and mortality are noted in the literature across the socioeconomic gradient (also termed socioeconomic status or SES) (ACS CAN, 2009; Harper and Lynch, 2005; Harper et al, 2008). These disparities are seen even in the case of lung cancer. Neighborhood deprivation, which represents an area-level measure of socioeconomic status, also shows consistent positive association with cancer incidence and mortality rates (Li,et al, 2012a; Li et al, 2012b; Sundquist et al, 2012; Zeigler-Johnson, 2011). This association persists even after controlling for individual-level socioeconomic variables, strengthening the notion of the environment as an important health exposure

Socioeconomic status, as commonly measured by education, occupation, employment, income and wealth, impacts every factor along the cancer continuum. Commonly hypothesiszed pathways include social, behavioral, and economic factors such as persistent inequalities in access to care, language barriers, unhealthy behaviors and unhealthy environments. While poor access to healthcare represents barriers to preventive screenings, delayed diagnoses and treatment, it is important to understand upstream connections between neighborhood disadvantage/low SES and unhealthy behaviors such as tobacco smoking, poor diets and physical inactivity.

SES shows a consistent, inverse relationship with unhealthy behaviors such as tobacco use, physical inactivity and poor diet. Pampel et al (2010), for example, show that high school dropouts have 3.7 times larger odds of smoking compared to college graduates. The odds remain larger even for lower occupation and income groups. Two significant mechanisms linking neighborhood deprivation and unhealthy behaviors include poor access to resources and chronic stress.

Individuals with lower SES have fewer resources to avail of health-supporting aids, lower levels of education leading to lack of health-related knowledge and information and lower levels of self-efficacy. This includes a lack of awareness with regard to health risks associated with tobacco smoking (Siahpush et al, 2006b). Collectively, disadvantaged neighborhoods have poor access to large grocery stores, making it harder to access healthy foods. Conversely, they have a higher concentration of liquor stores, fast-food restaurants and places to buy tobacco products. Research has shown that low SES neighborhoods can have equal amounts of gyms, parks and recreational facilities compared to high SES neighborhoods. However, utilization in lower SES neighborhoods maybe hampered due to perceived and real factors such as crime/safety and fewer desirable public places. It is also important to consider that this is a direct manifestation of social processes in the built environment. Even more important, it suggests the presence of land use nuances which may be difficult to detect with coarser land use data. In the absence of such fine-grained and qualitative data, SES or neighborhood deprivation indices might also be good proxy measures for land use.

Other direct social impacts include lack of access to employment opportunities and lower levels of social support and social cohesion. Social networks appear to have a stronger influence on smoking cessation among high SES groups compared to low SES groups (Christakis and Fowler, 2008). Social cohesion tends to be higher in high SES neighborhoods which has been shown to promote and support healthy behaviors including avoidance of smoking (Siahpush et al, 2006b).

Socioeconomically deprived individuals living in disadvantaged neighborhoods are subject to chronic stressors on a consistent basis. These stressors include higher unemployment, discrimination, higher crime, isolation, marginalization and powerlessness (Pampel et al, 2010). In these circumstances, unhealthy behaviors such as smoking, overeating and inactivity perform coping/self-medicating functions that become surrogates for pleasure (Lantz et al, 2005; Layte & Whelan, 2009).

The following sections present the results of the data analysis including land use pattern and socioeconomic variables.

Exploratory data analysis: Hierarchical clustering of lung cancer rates, land use indices and the neighborhood deprivation index

Hierarchical Clustering was used as a first step in exploring patterns in the data.. All tracts were plotted against their land use indices and Neighborhood Deprivation Index values. The values in each cell are assigned a value that range between red (high) to green (low). The branching patterns on the right (black) indicate the growth of clusers from finest to coarsest.

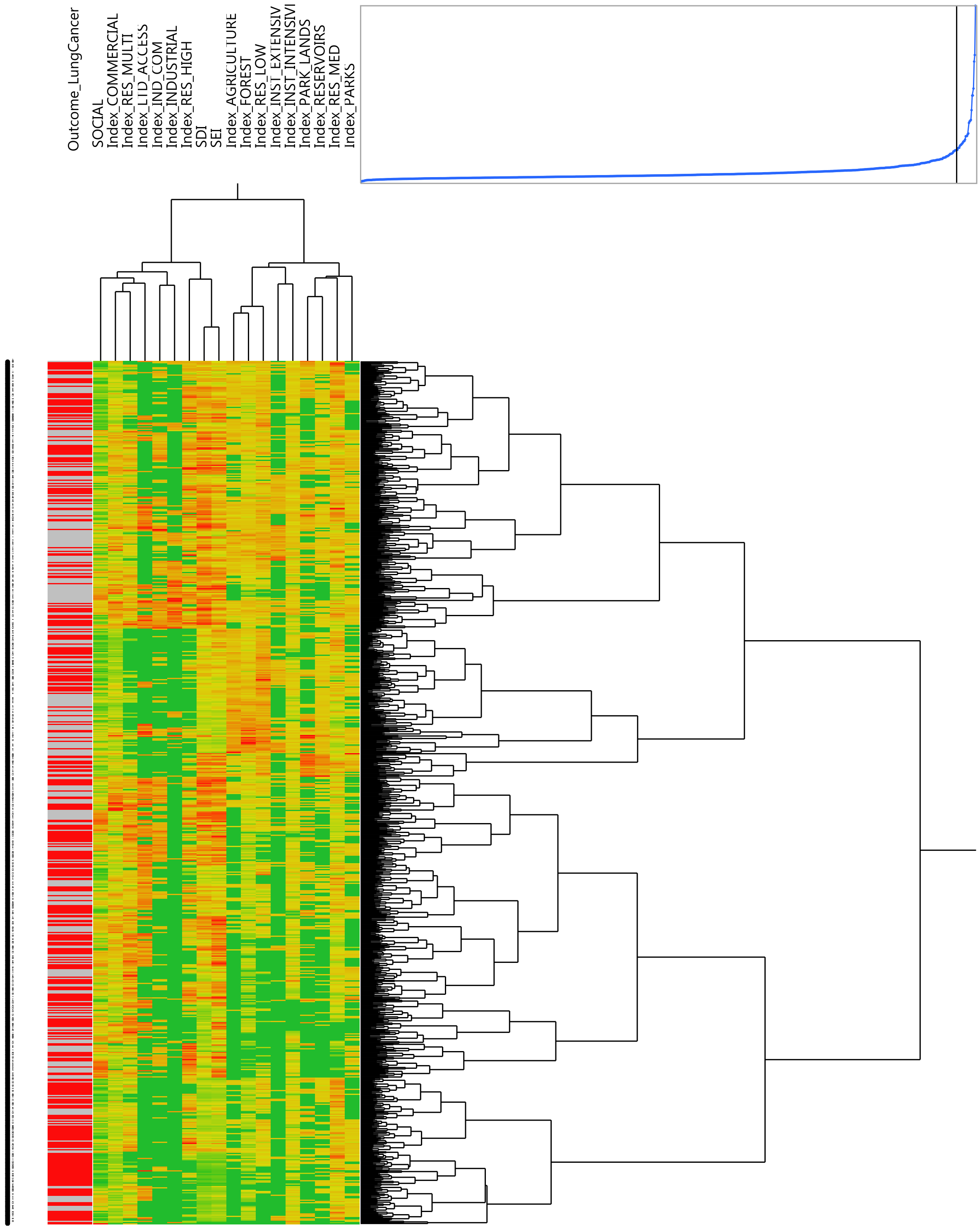


Figure 34. Hierarchical clustering of all 951 tracts with land use and socioeconomic variables

Figure 34 shows heirarchic clusterting applied to the 951 census tracts (as observations, represented by each horizontal line) and the clusterting variables as columns (comprising of SEI, SDI, SOCIAL and the land use indices). Two-way clustering is used to show the similarities in both row and columns space. When all 951 tracts are included, it is difficult to detect a discernible pattern that clearly distinguishes between tracts with high mortality rates vs. tracts with low mortality rates. JMP software used the CCC (Cubic Cluster Crirteria) as the method of assessing the presence of clusters in the dataset, and at 12 clusters, the value of CCC turns positive indicating that this is the minimum number of clusters that the algorithm has detected. This is also shown by the intercept of the blue line above.

For the 12 clusters that were detected, the next step was to assess the number of tracts that had high lung cancer. This is shown in table 9 below.

Table 9. 12 clusters identified with counts of tracts in high and low categories

	Outcome_LungCancer	
ClusterID_LungCancer	**1**	**0**
1	62	87
2	50	21
3	43	31
4	54	55
5	24	5
6	13	13
7	41	53
8	60	92
9	21	27
10	21	17
11	20	61
12	18	62

It is interestig to note that some clusters (E.g. #5) have 24 tracts with high lung cancer rates and only 5 that are low (high contrast). We also have clusters like #11 and

#12, where the number of high-risk tracts were 20 and 18 compared to the low-risk tract counts of 61 and 62 respectively. It is possible that the heirarchic clusterting method, used as part of an initial exploratory analysis, might have detected an interesting pattern that could merit further analysis.

Figure 35 below shows a detail of the heirarhical cluatering analysis done by selecting tracts in cluster 5 and 12, to highlight patterns in potential variables that might be related to the contrast in lung cancer rates between these clusters.

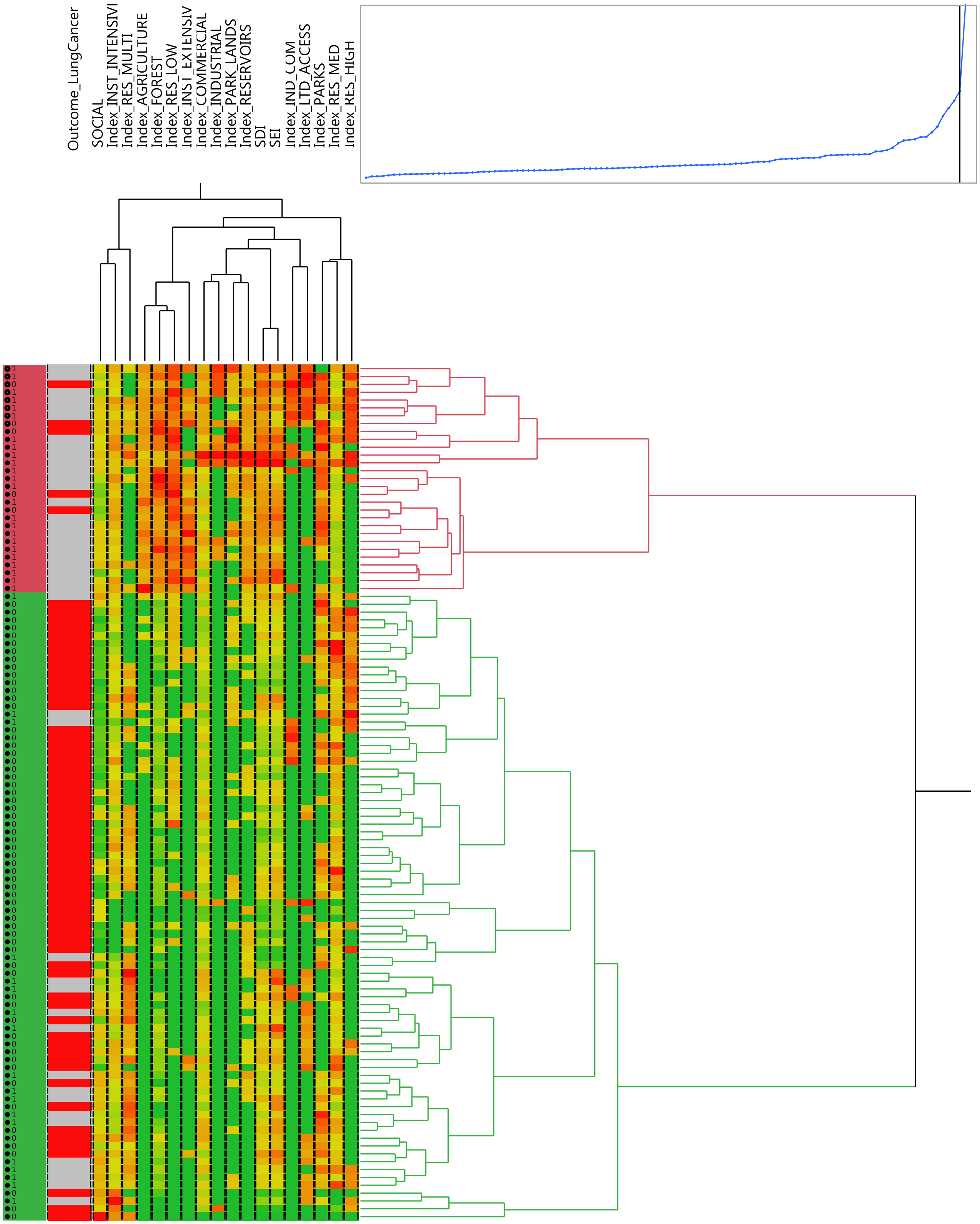


Figure 35. Hierarchical clustering of clusters 5 and 12

Here, the CCC metric suggested a minimum of 2 clusters, and we see that while these do not correspond exactly to tracts with low and high lung cancer eates, there is a more discernible pattern that exists. The grey rows indicate presence of high lung cancer rates in that tract. The relationship with the Neighborhood Deprivation Index (SOCIAL) is more concistsent where tracts with low values correspond consistently with low mortality rates. However, there also appears to be secondary pattern where the topmost cluster (predominantly grey) shows a high correspondnce of values associaed with these indices : Agriculture, Forest, Res_Low, and appear to be even more consistent than the SOCIAL index. This indicates that there may be an underlying pattern or combination of social and land use indices that could be associated with high lung cancer risks. Furthermore, it is also possible that land use indices have different levels of impact in different geographic locations indicating spatial heterogeneity. To investigate this further, more formal confirmatory modeling and data-mining methods are used. They include tree-based methods drawn from the data-mining field (like Random Forest) as well as more traditional regression-bassed methods (E.g. Logictic Regression, Stepwise Logistic Regression, Mixed-Effects Logistic Regression, etc.).

Level 1 Variable selection

Variable selection using level 1 land use indices (composite index of all land use metrics) were performed using the following methods:

1. Random Forest which is an exploratory technique that detects variable importance

2. Full Logistic regression that detects statistically significant relationships, as well as their direction and magnitude
3. Stepwise logistic regression that detects statistically significant relationships, as well as their direction and magnitude and also eleiminates non-significant variables.
4. Multilevel Logistic Regression which tests if tract level variables stay significant even after adding county level random effects. The model also detects if county level effects reduce/increase the impact of tract level on lung cancer mortality risk

Models 1 through 4 are run iteratively five times, each time with a new training and validation set. This methodology enables the detection of variables that remain consistently significant between methods and across iterations as well.

Level 1 Variable Importance: Random Forest

Table 10. Level 1 Variable Importance: Random Forest

Level 1 Indices	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Random Forest Variable Importance	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini
SDI	11.450	22.605	5.485	20.491	6.595	21.814	8.254	21.937	8.421	21.430
SEI	8.572	22.012	8.861	21.682	10.211	22.324	9.671	22.587	6.780	21.659
SOCIAL	45.698	68.449	47.421	67.756	48.181	65.730	42.668	63.368	47.833	64.400
Index_AGRICULTURE	14.376	16.921	17.164	18.682	17.940	19.448	14.654	19.511	18.286	20.381
Index_COMMERCIAL	6.361	21.132	3.111	20.376	2.855	21.107	2.877	22.731	0.690	19.577
Index_FOREST	12.003	26.006	14.755	27.971	14.215	28.274	16.051	30.226	16.083	31.137
Index_IND_COM	4.007	10.933	2.839	11.369	2.344	12.034	-0.339	10.023	1.647	10.170
Index_INDUSTRIAL	8.673	8.805	10.637	9.853	4.351	6.452	1.926	6.188	4.606	6.285
Index_INST_EXTENSIVE	3.320	7.532	2.769	6.087	3.132	7.214	7.227	9.221	5.670	8.528
Index_INST_INTENSIVE	0.213	18.247	1.415	20.034	1.153	19.488	0.093	18.238	1.366	19.534
Index_LTD_ACCESS	1.923	10.110	3.594	10.574	4.140	10.415	1.348	10.238	2.691	10.252
Index_PARK_LANDS	1.994	8.359	-0.036	8.976	-0.980	9.524	2.951	9.749	-0.150	8.540
Index_PARKS	3.625	13.636	1.519	13.510	1.781	15.503	3.162	13.965	0.998	13.519
Index_RES_LOW	8.854	19.697	9.728	20.044	6.499	18.344	6.391	18.125	10.174	20.369
Index_RESERVOIRS	3.050	13.407	2.761	13.361	0.254	12.669	-0.340	13.156	4.703	13.911
Index_RES_HIGH	5.555	17.000	4.103	15.670	5.247	16.717	4.705	17.652	7.052	15.773
Index_RES_MED	4.810	24.055	6.958	23.441	4.813	21.982	2.270	22.560	4.600	22.468
Index_RES_MULTI	9.649	23.535	10.834	22.998	10.699	23.561	8.599	23.306	14.002	24.687
Confusion Matrix, Predicted	0	1	0	1	0	1	0	1	0	1
Training, Actual 0	303	90	302	91	295	98	298	95	297	96
1	108	213	105	216	109	212	105	216	104	217
Testing, Actual 0	93	38	93	38	105	26	104	27	96	35
1	33	73	41	65	33	73	37	69	38	68
Lung Cancer	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Classification Accuracy	72.3%	70.0%	72.5%	66.7%	71.0%	75.1%	72.0%	73.0%	72.0%	69.2%

The Neighborhood Deprivation Index (SOCIAL) emerges consistently across models as the most significant variable. Next to the SOCIAL, the Agriculture index is the second most important variable as denoted by the “Mean Decrease Accuracy” and “Mean Decrease Gini”. A variable that causes a large decrease in the accuracy of a Random Forest indicates that it is an improtant variable for the classification of data (it creates a large decerase in accuracy when excluded). A higher decrease in Gini means that a particular predictor variable plays a greater role in partitioning the data into the defined classes. In other words, both measures of variable importance indicate how much more helpful than random a particular predictor variable is in successfully classifying data. However, it is important to note that the SOCIAL variable cause a much more dramatic drop in importance measures while the change in importance measure are much more subtle with the Land Use variables. This largely indicates that Land Use indices are a secondary effect.

Level 1 Confirmatory Data Analysis

Logistic Regression with Level 1 Land Use Indices

Table 11. Results of Logistic Regression with Level 1 Land Use Indices

FULL LOGISTIC REGRESSION	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	-0.50302	0.40894	-0.56501	0.35220	-1.17682	0.04418	-1.06799	0.08717	-1.08751	0.08312
SDI	-2.46972	0.03722	-1.74242	0.12855	-0.67723	0.53621	-1.91797	0.09483	-1.84176	0.12105
SEI	5.94354	0.01620	4.35454	0.06457	2.93634	0.19868	5.51561	0.02283	5.39311	0.03021
SOCIAL	0.47593	0.00000	0.48140	0.00000	0.45264	0.00000	0.47398	0.00000	0.49759	0.00000
Index_AGRICULTURE	0.20470	0.00014	0.22812	0.00002	0.20627	0.00008	0.17416	0.00117	0.20419	0.00016
Index_COMMERCIAL	0.11881	0.01630	0.10890	0.03342	0.13163	0.00934	0.16397	0.00213	0.11155	0.02780
Index_FOREST	0.00670	0.89333	0.02616	0.60237	0.01908	0.71571	0.07643	0.15279	-0.00288	0.95620
Index_IND_COM	0.00796	0.83207	-0.01186	0.75630	0.00744	0.83769	-0.01391	0.70936	0.01802	0.64141
Index_INDUSTRIAL	0.10835	0.00412	0.09950	0.00294	0.03803	0.22472	0.04676	0.16854	0.08743	0.01169
Index_INST_EXTENSIVE	0.03467	0.28843	0.02426	0.48909	0.03693	0.24983	0.04669	0.15742	0.07043	0.03969
Index_INST_INTENSIVE	0.00465	0.91561	-0.05706	0.17724	-0.10328	0.02739	-0.02733	0.53370	-0.05737	0.19183
Index_LTD_ACCESS	0.02768	0.43109	0.02082	0.53994	-0.01853	0.58495	0.01438	0.67765	0.01893	0.59024
Index_PARK_LANDS	0.03380	0.41382	0.01451	0.72098	-0.00370	0.92420	-0.01309	0.75176	0.00329	0.93666
Index_PARKS	0.02284	0.53272	0.02989	0.42792	0.01242	0.74163	0.02398	0.50637	0.02921	0.43003
Index_RES_LOW	0.02286	0.67947	-0.02473	0.65925	-0.01132	0.83130	0.01136	0.84129	0.04697	0.40997
Index_RESERVOIRS	0.01688	0.68746	-0.01236	0.75059	-0.02584	0.47961	-0.01035	0.78237	0.00233	0.95026
Index_RES_HIGH	0.03483	0.37234	0.03035	0.42717	-0.00322	0.93487	0.03990	0.29894	0.01376	0.72696
Index_RES_MED	0.22879	0.00003	0.23853	0.00002	0.19829	0.00025	0.17833	0.00129	0.20689	0.00027
Index_RES_MULTI	-0.09874	0.03142	-0.10220	0.02340	-0.13014	0.00395	-0.13668	0.00311	-0.14509	0.00204
Confusion Matrix, Predicted	0	1	0	1	0	1	0	1	0	1
Training, Actual 0	310	83	308	85	311	82	312	81	311	82
1	117	204	118	203	118	203	112	209	110	211
Testing, Actual 0	92	39	103	28	100	31	100	31	98	33
1	36	70	44	62	42	64	49	57	41	65
Lung Cancer	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Classification Accuracy	72.0%	68.4%	71.6%	69.6%	72.0%	69.2%	73.0%	66.2%	73.1%	68.8%

Significant variables are highlighted in blue. Next to Social, Agricultural Index is consistently significant with low p-value and higher estimate compared to other land use variables.

Stepwise Logistic Regression with Level 1 Land Use Indices

Table 12. Results of Stepwise Logistic Regression with Level 1 Land Use Indices

STEPWISE LOGISTIC	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	-0.17851	0.04188	-0.23212	0.00734	-1.09761	0.03838	-1.04723	0.05756	-1.32580	0.01768
SOCIAL	0.47893	0.00000	0.48862	0.00000	0.46633	0.00000	0.48673	0.00000	0.49790	0.00000
Index_AGRICULTURE	0.18783	0.00000	0.19911	0.00000	0.21020	0.00000	0.19230	0.00000	0.19327	0.00000
Index_RES_MED	0.17076	0.00014	0.18644	0.00005	0.17759	0.00013	0.12810	0.00524	0.16612	0.00061
Index_INDUSTRIAL	0.08745	0.01219	0.07730	0.01139					0.07612	0.01944
Index_RES_MULTI	-0.10901	0.00913	-0.10087	0.01413	-0.12973	0.00249	-0.15210	0.00050	-0.16516	0.00020
Index_COMMERCIAL	0.10946	0.01627	0.10975	0.01997	0.13080	0.00503	0.15534	0.00108	0.09918	0.04015
SEI					1.32803	0.10109	1.26010	0.13145	1.69407	0.04546
Index_INST_EXTENSIVE									0.05576	0.07955
Index_INST_INTENSIVE			-0.07055	0.06768	-0.09932	0.02221			-0.07333	0.07892
Confusion Matrix, Predicted	0	1	0	1	0	1	0	1	0	1
Training, Actual 0	317	76	311	82	310	83	304	89	312	81
1	119	202	118	203	114	207	113	208	105	216
Testing, Actual 0	93	38	106	25	95	36	103	28	100	31
1	37	69	43	63	38	68	42	64	43	63
Lung Cancer	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Classification Accuracy	72.7%	68.4%	72.0%	71.3%	72.4%	68.8%	71.7%	70.5%	73.9%	68.8%

Significant variables are highlighted in blue. Next to Social, Agricultural Index is significant as indicated by low p-value and has the highest so-efficient estimate among all other land use indices. These values also stay consistent across all five iterations. Classification accuracy consistently hovers around 70%, indicating that 70% of tracts are correctly classified for each iteration. The indices for Medium Density Residential, Multi Family Residential and Commercial land uses are also significant. However, their effect sizes are much smaller as indicated by the z scores corresponding with their p-values. It is interesting to note an inverse relationship between mortality rates and Multi Family Residential.

Multilevel Logistic Regression with Level 1 Land Use Indices

Table 13. Results of Multilevel Logistic Regression with Level 1 Land Use Indices

MULTILEVEL LOGISTIC	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	0.07710	0.71392	-0.13139	0.40031	-1.11415	0.05531	-1.03205	0.07870	-1.32436	0.02300
SOCIAL	0.45532	0.00000	0.47777	0.00000	0.45415	0.00000	0.47695	0.00000	0.49173	0.00000
Index_AGRICULTURE	0.14596	0.00122	0.17853	0.00003	0.17005	0.00016	0.16911	0.00007	0.17835	0.00002
Index_RES_MED	0.12036	0.01264	0.16506	0.00101	0.13917	0.00469	0.11147	0.02115	0.15197	0.00279
Index_INDUSTRIAL	0.09238	0.01065	0.07649	0.01332					0.07300	0.02640
Index_RES_MULTI	-0.10224	0.01926	-0.09581	0.02401	-0.13551	0.00282	-0.14935	0.00087	-0.16482	0.00028
Index_COMMERCIAL	0.10983	0.02276	0.11248	0.02126	0.12981	0.00755	0.15250	0.00174	0.10114	0.04130
SEI					1.70376	0.04516	1.43441	0.10020	1.80977	0.03822
Index_INST_EXTENSIVE									0.06005	0.06993
Index_INST_INTENSIVE			-0.06337	0.11057	-0.09488	0.03561			-0.06949	0.10229
Random Effects - County	Estimate		Estimate		Estimate		Estimate		Estimate	
Barrow	0.68375		0.30778		0.59539		0.22127		0.15756	
Bartow	0.26943		0.17968		0.35671		0.15076		0.08380	
Carroll	0.20314		0.01787		0.09773		0.03127		-0.01602	
Cherokee	-0.19452		0.00069		-0.01533		0.02019		-0.02901	
Clayton	0.48063		0.17262		0.17778		0.22465		0.19294	
Cobb	-0.16383		-0.08234		-0.20525		-0.10533		0.05882	
Coweta	-0.26991		-0.02061		0.12139		0.08654		0.05958	
Dawson	0.13092		-0.03865		0.14497		0.03508		-0.01939	
DeKalb	-0.30152		-0.04215		-0.02473		-0.10987		-0.07018	
Douglas	0.14781		0.10724		0.08480		0.15917		0.10666	
Fayette	-1.12482		-0.35659		-0.63029		-0.28134		-0.17645	
Forsyth	-0.35119		-0.05659		-0.42155		-0.17150		-0.15598	
Fulton	-0.77196		-0.36663		-0.76185		-0.40495		-0.27718	
Gwinnett	-0.43791		-0.24934		-0.39358		-0.26579		-0.30157	
Hall	-0.08252		-0.07531		-0.19715		-0.06980		-0.08114	
Henry	0.51113		0.12927		0.47201		0.12490		0.11162	
Newton	0.17189		0.03727		0.04421		0.12150		0.07402	
Paulding	0.80391		0.25147		0.39072		0.19280		0.19933	
Rockdale	-0.00571		-0.01984		-0.14432		0.00384		0.01366	
Spalding	0.45344		0.12132		0.28272		0.06492		0.07531	
Walton	-0.29518		-0.02434		-0.05787		-0.03965		-0.01088	
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual 0	357	36	354	39	352	41	355	38	352	41
1	158	163	171	150	161	160	167	154	160	161
Testing, Actual 0	103	28	118	13	116	15	116	15	113	18
1	51	55	59	47	55	51	56	50	55	51
Lung Cancer	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy	**72.8%**	**66.7%**	**70.6%**	**69.6%**	**71.7%**	**70.5%**	**71.3%**	**70.0%**	**71.8%**	**69.2%**

In the multilevel model, tract variables stay significant even after adding county level random effects (random intercept, fixed slopes). However, it is interesting to note that the estimates of the intercepts for each county vary. Positive estimates indicate that there are county-level effects that increase the risk of lung cancer mortality within tracts. Conversely, negative estimates indicate that county-level effects ameliorate the lung cancer mortality risk within tracts. The counties with negative intercepts are consonant with the healthy county groupings identified in Chapter III in the section on Hierarchical Clustering (pg.65) as well as figure 44 in Chapter VI.

Level 2 Variable Selection

The Level 1 analysis provided insights into which Land Uses were important. The Level 2 analysis is meant to analyze which aspect (Geometry, Shape or Interspersion) of the Land Uses bear a significant relationship with mortality rates. Variable selection using level 2 land use indices (disaggregated indices based on landuse metrics grouped into Geometry, Shape and Inetrspersion domains) were performed using the following methods:

1. Random Forest which is an exploratory technique that detects variable importance
2. Full Logistic regression that detects statistically significant relationships, as well as their direction and magnitude
3. Stepwise logistic regression that detects statistically significant relationships, as well as their direction and magnitude and also eleiminates non-significant variables.

Level 2 Variable Importance: Random Forest

Table 14. Level 2 Variable Importance: Random Forest

RANDOM FOREST, L2	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Importance	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini
SOCIAL	49.345	76.802	46.562	75.324	51.867	78.615	44.906	73.025	44.249	73.172
IDX_GEOM_AGRICULTURE	11.159	16.647	13.642	17.286	17.199	19.509	17.097	19.002	15.024	17.237
IDX_SHAPE_AGRICULTURE	9.594	14.414	12.100	17.888	11.811	15.626	13.300	16.666	11.347	16.097
IDX_INTER_AGRICULTURE	9.023	13.645	9.633	13.497	10.993	13.637	7.293	13.685	6.374	13.256
IDX_GEOM_COMMERCIAL	2.240	22.716	0.318	22.715	1.821	22.826	1.365	21.313	3.203	22.687
IDX_SHAPE_COMMERCIAL	-0.777	23.573	3.013	22.597	2.219	23.679	2.951	22.857	4.335	24.607
IDX_INTER_COMMERCIAL	0.364	22.842	-0.765	20.468	4.385	23.028	3.603	21.439	1.316	22.396
IDX_GEOM_INDUSTRIAL	5.073	5.998	4.391	6.070	3.555	5.056	5.903	6.070	4.328	5.125
IDX_SHAPE_INDUSTRIAL	3.361	6.391	1.696	5.243	0.782	4.180	4.839	6.142	4.321	4.633
IDX_INTER_INDUSTRIAL	2.707	5.680	3.290	4.172	1.417	4.507	4.917	5.774	3.295	4.156
IDX_GEOM_RES_MED	10.503	31.273	12.067	31.988	11.249	31.135	12.145	30.827	14.993	32.006
IDX_SHAPE_RES_MED	-0.913	22.379	4.277	24.091	3.535	22.473	2.878	23.015	0.081	22.625
IDX_INTER_RES_MED	1.367	24.130	4.053	24.783	4.230	23.881	1.183	22.277	4.374	26.718
IDX_GEOM_RES_MULTI	9.928	25.301	16.494	27.580	14.061	25.254	14.408	27.498	19.213	28.559
IDX_SHAPE_RES_MULTI	3.106	19.185	2.366	17.833	3.491	18.595	3.824	20.670	1.422	18.897
IDX_INTER_RES_MULTI	4.956	21.038	6.663	20.849	7.827	20.328	8.987	22.076	2.692	19.972
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	300	93	300	93	304	89	299	94	304	89
1	122	199	107	214	111	210	114	207	108	213
Testing, Actual **0**	102	29	98	33	97	34	95	36	92	39
1	34	72	43	63	45	61	39	67	38	68
Lung Cancer	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy	**69.9%**	**73.4%**	**72.0%**	**67.9%**	**72.0%**	**66.7%**	**70.9%**	**68.4%**	**72.4%**	**67.5%**

The Geometry Index (GEOM) for Agriculture, Medium Density Residential and MultiFamily Resdiential are the most important variables as denoted by the Mean Decrease Accuracy and Mean Decerase Gini scores.

Level 2 Confirmatory Data Analysis: Logistic Regression with Level 2 Land Use Indices

Table 15. Results of Logistic Regression with Level 2 Land Use Indices

FULL LOGISTIC REGRESSION, L2	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	-0.23013	0.00823	-0.24078	0.00652	-0.20811	0.01835	-0.19502	0.02498	-0.23487	0.00766
SOCIAL	0.44738	0.00000	0.50325	0.00000	0.48114	0.00000	0.43299	0.00000	0.44775	0.00000
IDX_GEOM_AGRICULTURE	0.22148	0.00161	0.19307	0.00725	0.23924	0.00066	0.22764	0.00172	0.22830	0.00172
IDX_SHAPE_AGRICULTURE	-0.00607	0.93648	-0.01184	0.87982	0.00500	0.94783	0.00842	0.91340	-0.03176	0.67930
IDX_INTER_AGRICULTURE	0.03698	0.72174	0.11751	0.29372	0.09243	0.40267	0.06422	0.54075	0.10078	0.32955
IDX_GEOM_COMMERCIAL	0.11577	0.12966	0.14998	0.05637	0.15556	0.05236	0.06464	0.37982	0.10364	0.17039
IDX_SHAPE_COMMERCIAL	-0.03000	0.60636	-0.00677	0.91565	0.08128	0.18365	-0.00682	0.90939	0.02426	0.73356
IDX_INTER_COMMERCIAL	0.01134	0.91397	-0.00253	0.98146	-0.09454	0.36994	-0.03206	0.75237	-0.01530	0.88142
IDX_GEOM_INDUSTRIAL	0.05094	0.51148	0.02265	0.76269	0.06837	0.37952	0.04562	0.57845	0.05836	0.45997
IDX_SHAPE_INDUSTRIAL	0.03477	0.52632	0.03410	0.55845	-0.00470	0.93445	0.04215	0.47250	0.03503	0.52607
IDX_INTER_INDUSTRIAL	0.04325	0.73994	-0.00523	0.95867	0.03251	0.79506	0.01598	0.88335	-0.06529	0.54764
IDX_GEOM_RES_MED	0.20174	0.00726	0.17650	0.02078	0.20018	0.00766	0.22481	0.00316	0.22773	0.00253
IDX_SHAPE_RES_MED	0.02858	0.51543	0.32864	0.00966	0.03189	0.46936	0.01177	0.80336	0.02370	0.59527
IDX_INTER_RES_MED	0.20369	0.02048	0.14222	0.13342	0.26728	0.00263	0.18904	0.03206	0.22720	0.00995
IDX_GEOM_RES_MULTI	-0.14456	0.05106	-0.30488	0.00010	-0.20637	0.00653	-0.20365	0.00679	-0.28295	0.00036
IDX_SHAPE_RES_MULTI	0.07268	0.23594	0.07410	0.20884	0.06446	0.28312	0.12461	0.03327	0.07379	0.21064
IDX_INTER_RES_MULTI	-0.13050	0.25543	-0.06073	0.59363	-0.03287	0.78053	-0.08793	0.44720	-0.05959	0.59283
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	316	77	316	77	317	76	309	84	310	83
1	126	195	110	211	114	207	112	209	116	205
Testing, Actual **0**	106	25	99	32	95	36	97	34	97	34
1	42	64	44	62	42	64	38	68	35	71
Lung Cancer	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy	**71.6%**	**71.7%**	**73.8%**	**67.9%**	**73.4%**	**67.1%**	**72.5%**	**69.6%**	**72.1%**	**70.9%**

Significant variables shown in blue. Again Geometry Index for Agriculture emerges the most significant of all the Land Use Indices next to SOCIAL.

Stepwise Logistic Regression with Level 2 Land Use Indices

Table 16. Results of Stepwise Logistic Regression with Level 2 Land Use Indices

STEPWISE LOGISTIC, L2	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	-0.23280	0.00735	-0.24023	0.00632	-0.20770	0.01774	-0.18968	0.02821	-0.23248	0.00790
SOCIAL	0.44037	0.00000	0.50093	0.00000	0.46306	0.00000	0.42551	0.00000	0.44567	0.00000
IDX_GEOM_AGRICULTURE	0.22156	0.00016	0.19534	0.00124	0.24273	0.00003	0.23769	0.00015	0.21037	0.00060
IDX_GEOM_RES_MED	0.20329	0.00152	0.21269	0.00129	0.21305	0.00086	0.24417	0.00018	0.24294	0.00015
IDX_GEOM_RES_MULTI	-0.16405	0.00521	-0.30951	0.00000	-0.20710	0.00044	-0.22658	0.00020	-0.28881	0.00000
IDX_GEOM_INDUSTRIAL	0.09831	0.04606			0.08295	0.08407				
IDX_INTER_RES_MED	0.21501	0.01108	0.14966	0.10618	0.27985	0.00127	0.18343	0.03362	0.23166	0.00733
IDX_GEOM_COMMERCIAL	0.12844	0.03372	0.17174	0.00700	0.14009	0.02934			0.12276	0.03834
IDX_SHAPE_RES_MED			0.35071	0.00538						
IDX_SHAPE_INDUSTRIAL							0.07973	0.04997		
IDX_SHAPE_RES_MULTI							0.11716	0.03145		
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	324	69	312	81	320	73	313	80	308	85
1	131	190	111	210	123	198	118	203	119	202
Testing, Actual **0**	107	24	94	37	98	33	97	34	98	33
1	46	60	44	62	42	64	36	70	39	67
Lung Cancer	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy	**72.0%**	**70.5%**	**73.1%**	**65.8%**	**72.5%**	**68.4%**	**72.3%**	**70.5%**	**71.4%**	**69.6%**

Significant variables shown in blue. Both Logistic models confirm that the Geometry Index for Agriculture, Medium Density Residential and MultiFamily Residential are consistently significant between iterations and across models. MultiFamily residential maintains its negative association with lung cancer mortality rates. The Geometry Index for Agriculture has the largest coefficient estimate as indicate by the estimate and associated p values.

Multilevel Logistic Regression with Level 2 Land Use Indices

Table 17. Results of Multilevel Logistic Regression with Level 2 Land Use Indices

MULTILEVEL LOGISTIC , L2	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	0.03893	0.84440	-0.12265	0.40367	-0.04084	0.80662	0.04545	0.80660	-0.10097	0.52924
SOCIAL	0.43655	0.00000	0.51106	0.00000	0.47433	0.00000	0.43306	0.00000	0.45546	0.00000
IDX_GEOM_AGRICULTURE	0.15919	0.01700	0.16146	0.01283	0.19971	0.00200	0.18131	0.00897	0.18164	0.00566
IDX_GEOM_RES_MED	0.15994	0.02730	0.21590	0.00212	0.19855	0.00413	0.22663	0.00141	0.23504	0.00072
IDX_GEOM_RES_MULTI	-0.15117	0.01240	-0.30386	0.00000	-0.20815	0.00055	-0.22554	0.00028	-0.29122	0.00000
IDX_GEOM_INDUSTRIAL	0.10292	0.04220			0.08100	0.10092				
IDX_INTER_RES_MED	0.17033	0.05530	0.12836	0.17675	0.21565	0.01927	0.11365	0.21446	0.18237	0.04485
IDX_GEOM_COMMERCIAL	0.12952	0.03870	0.16928	0.00971	0.15034	0.02398			0.13303	0.02866
IDX_SHAPE_RES_MED			0.30759	0.01535						
IDX_SHAPE_INDUSTRIAL							0.07486	0.07310		
IDX_SHAPE_RES_MULTI							0.13582	0.01579		
Random Effects - County	Estimate		Estimate		Estimate		Estimate		Estimate	
Barrow	0.43465		0.22217		0.36363		0.44120		0.24102	
Bartow	0.50097		0.17938		0.36954		0.45416		0.19759	
Carroll	0.16218		-0.09915		-0.06018		0.02351		0.05448	
Cherokee	-0.02120		0.16628		-0.11054		-0.06308		0.09781	
Clayton	0.33232		0.34504		0.25286		0.27141		0.31323	
Cobb	-0.31556		0.18138		0.11403		0.01037		-0.03129	
Coweta	0.05890		0.07789		-0.01986		0.01250		0.07024	
Dawson	-0.19100		-0.05198		-0.13919		0.06828		0.04939	
DeKalb	-0.41461		-0.11166		-0.13661		-0.29896		-0.04451	
Douglas	0.21130		0.03692		0.22923		0.30748		0.19507	
Fayette	-0.68595		-0.33025		-0.45866		-0.47445		-0.46346	
Forsyth	-0.39354		-0.14932		-0.16085		-0.52212		-0.25562	
Fulton	-0.76869		-0.55926		-0.74943		-0.91167		-0.62106	
Gwinnett	-0.45075		-0.35232		-0.33152		-0.41910		-0.20967	
Hall	-0.03447		0.00704		-0.18170		0.00197		-0.26003	
Henry	0.35147		0.00902		0.01753		0.05720		0.21347	
Newton	0.23730		0.04353		0.18015		0.20662		0.12897	
Paulding	0.57744		0.23385		0.27928		0.40178		0.49358	
Rockdale	0.07159		0.03099		0.22019		0.13324		-0.06439	
Spalding	0.17865		0.14344		0.23289		0.22264		0.07702	
Walton	0.08441		-0.04064		0.05131		0.00384		-0.21952	
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	358	35	362	31	353	40	352	41	352	41
1	160	161	153	168	162	159	160	161	159	162
Testing, Actual **0**	111	20	113	18	113	18	114	17	113	18
1	60	46	61	45	56	50	54	52	53	53
Lung Cancer	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy	**72.7%**	**66.2%**	**74.2%**	**66.7%**	**71.7%**	**68.8%**	**71.8%**	**70.0%**	**72.0%**	**70.0%**

Geometry Index for Agriculture, Medium Density Residential and MultiFamily Residential maintain their significance and direction at the tract level. Similar to the Level 1 analysis, Cherokee, Cobb, Dekalb, Fayette, Forsyth, Fulton, Gwinnett and Hall counties have negative intercept values, implying county-level characteristics that have an ameliorating effect on tract-level outcomes. Coweta, Dawson, Rockdale and Walton switch direction from negative to positive.

Analysis of Residuals

Pearson residuals are commonly used diagnostic measures in logistic regression. They are comparable to standardized residuals used in linear regression models. The Pearson residual for each tract equals the difference between observed and fitted probability for the tract, divided by an estimate of the standard deviation of the observed value. Mathematically, the Pearson residual for a given tract is,

$$Residual = \frac{Y - p}{\sqrt{p(1-p)}}$$

Where:
Y = Actual Outcome (0 or 1)
p = Predicted probability of moving to a "1" state, from the logistic or multilevel logistic model, which is a function of the significant variables

Pearson Residuals from a single iteration of four different lung cancer models were generated and mapped. The purpose of the residual analysis was threefold:

1. Identifying geographical locations where tracts are misclassified
2. Detection of spatial autocorrelation among residuals
3. Identifying outliers

Figure 36 shows the range of residuals for models using significant Level 1 indices. 4 tracts in each of the Level1 models have residuals less than -3. Two of these are less than -5. These tracts are very small in size and are located in the lower central part of the study area.

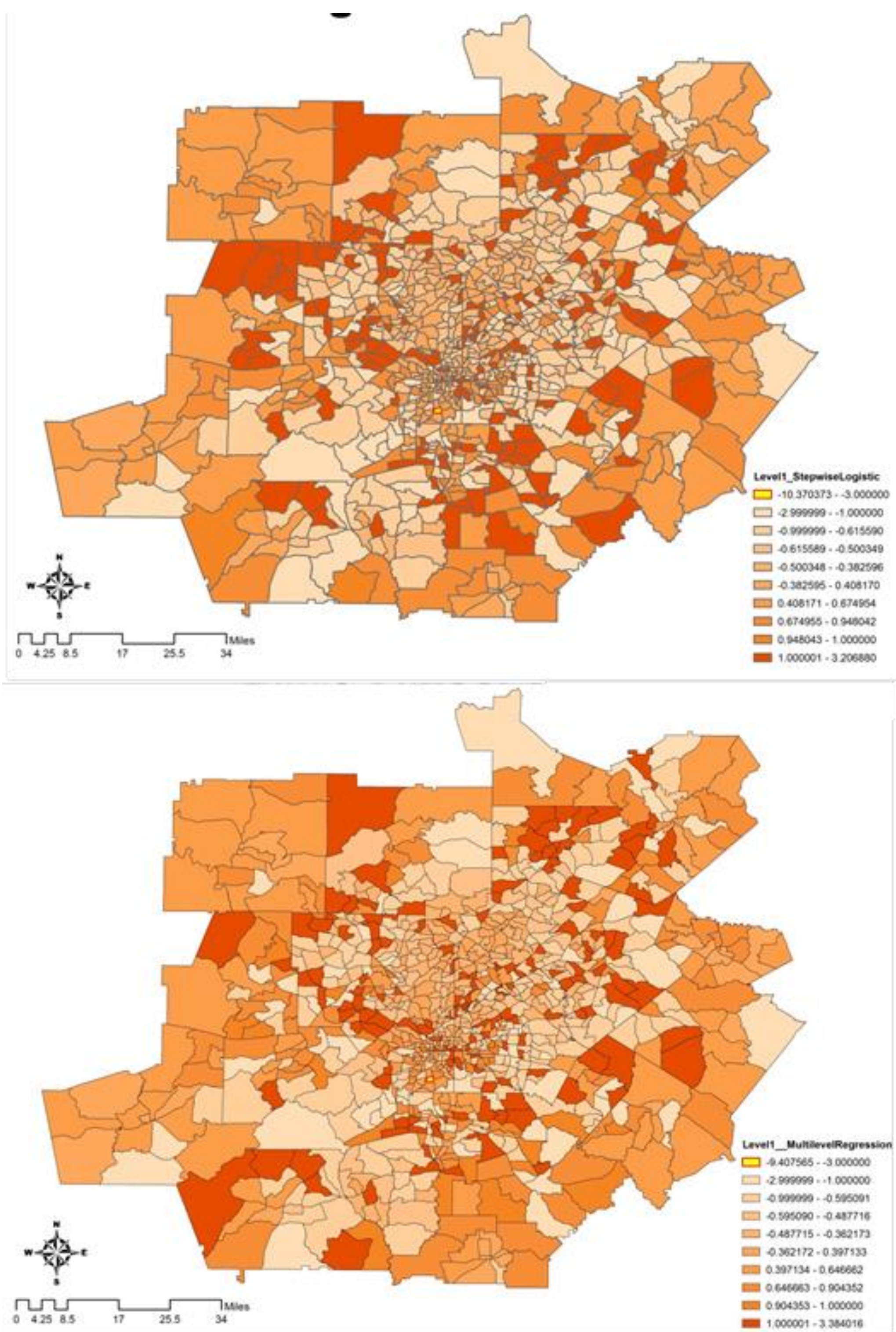
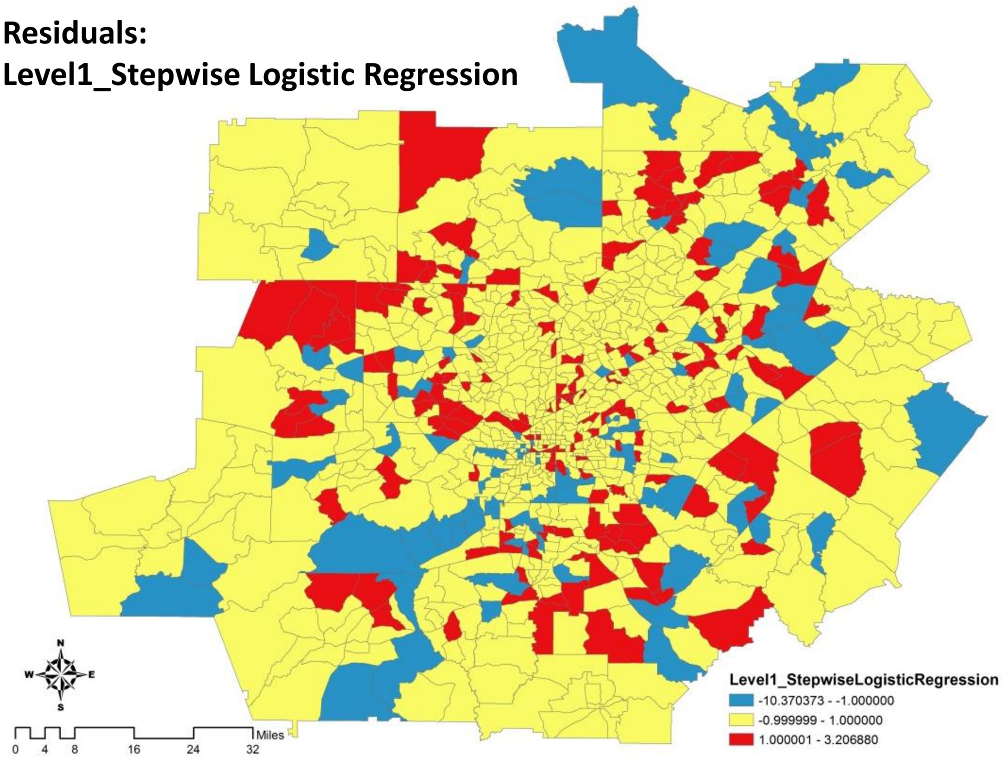


Figure 36. Pearson Residuals for Lung Cancer models calculated using Level 1 land use indices

Figures 37 and 38 are maps showing the misclassified tracts for each of the four scenarios. The blue tracts represent healthy (0) tracts misclassified as unhealthy (1) and the red tracts represent unhealthy tracts (1) misclassified as healthy (0). Predicted probabilities were calculated form the coefficients of the significant variables. 0.5 was chosen as the cutoff threshold to assign predicted membership in healthy (0) tracts or unhealthy (1) tracts for Lung Cancer. This was more stringent than the .45 cutoff based on the Youden Index derived from the ROC curve (Vazquez-Prokopec et al, 2012).

Table 18 provides a summary of the Moran's I statistics for each of the scenarios. It also provides a tally of the number of outliers (>3 and <-3) for each scenario. **The Moran's I statistic stayed consistently insignificant through all scenarios, indicating that there was no spatial autocorrelation among residuals.** The total number of outliers varied between 0.7% to 1.5% of the entire dataset. Models with Level 1 indices produced fewer outliers than models with Level 2 indices.



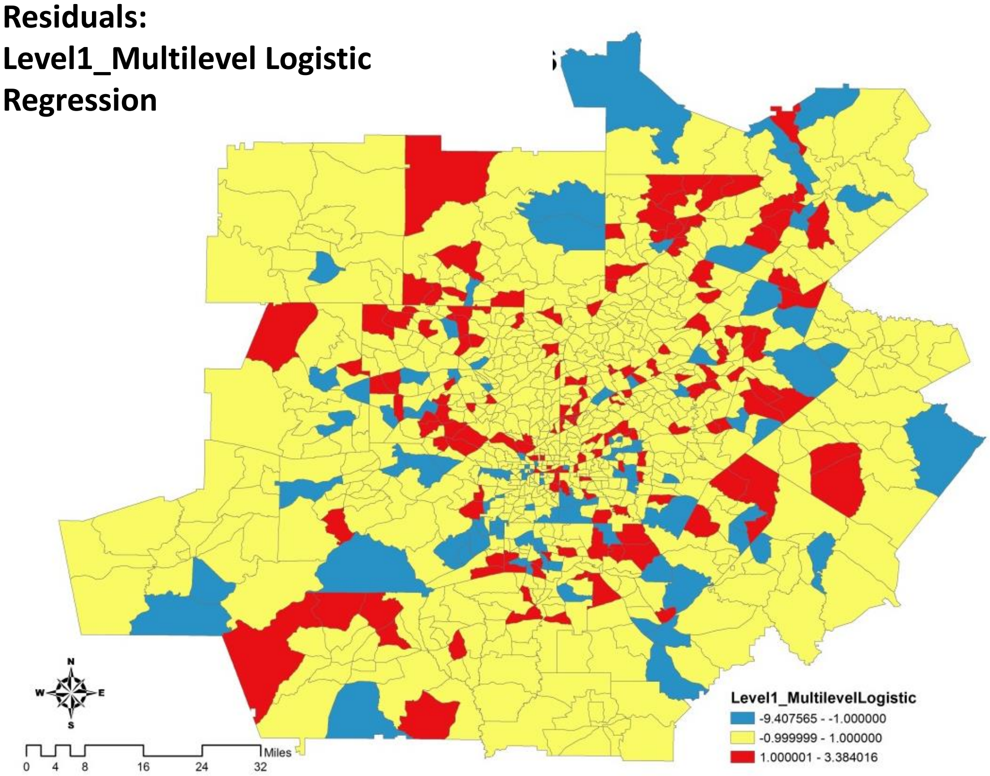
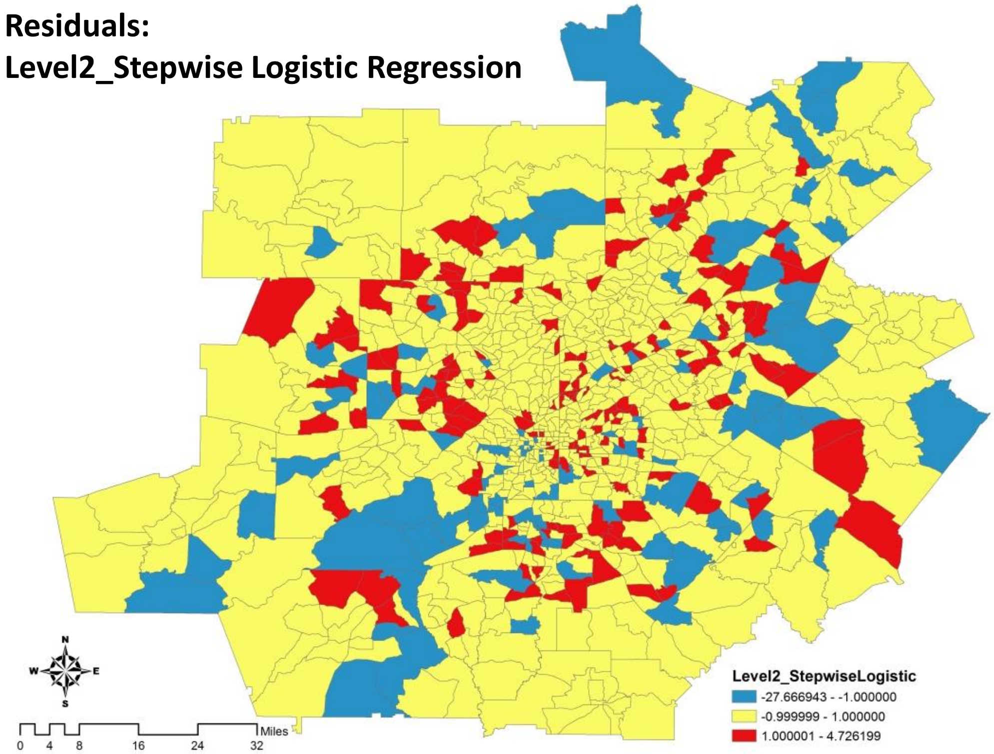


Figure 37. Residuals for Lung Cancer models using Level 1 Land Use Indices



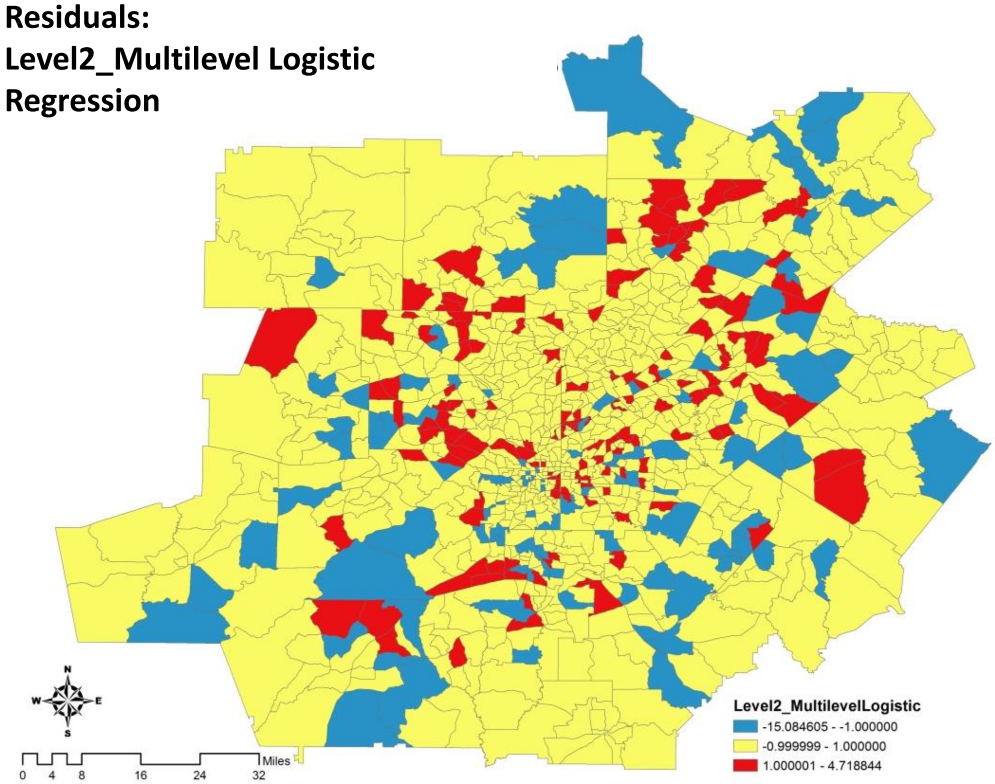


Figure 38. Residuals for Lung Cancer models using Level 2 Land Use Indices

Table 18. Table summarizing Moran's I statistics and outliers

Model	Moran's I	z-score	p-value	# Residuals < -3	# Residuals > +3
Level1_Stepwise Logistic Regression	0.025202	1.444	0.148727	4	3
Level1_ Multilevel Logistic Regression	0.011161	0.670670	0.502431	4	4
Level2_Stepwise Logistic Regression	0.020728	1.293591	0.195807	7	8
Level2_Multilevel Logistic Regression	0.012515	0.755729	0.449812	7	8

Discussion

Analytical Summary

From the analysis, the Neighborhood Deprivation (SOCIAL) Index shows the strongest association with Lung Cancer mortality rates, making it the primary effect. However, a secondary but statistically significant effect is seen in the case of Agriculture as a land use. It also has the largest co-efficient estimate and p-values among all the other significant land use variables. The Agriculture level 1 index shows a consistent, positive relationship with lung cancer mortality rates. Among the level 2 indices, the geometry index for agriculture again shows consistent associations with lung cancer mortality rates.

There is statistical evidence from multiple models that land use variables have an impact on lung cancer mortality risk, even after accounting for socioeconomic variables, adjusting for population age distributions and including county-level effects. Agriculture indices are consistently significant across all models and methods.

The models using the disaggregated indices show that the “GEOMETRY” Index is the most significant landscape metric compared to other landscape pattern measures. All landscape pattern metrics included in this category load prominently on the first Principal Component. It is challenging to sort the metrics strictly into composition vs configuration metrics. Furthermore, it is less relevant to interpret the metrics in isolation (hence indices are better). However, the metrics with largest loadings to the index include the Patch Size co-efficient of variation (PSCoV) and Edge Density. This indicates that configuration metrics are important, that spatial arrangement is significant. The Interspersion Juxtaposition Index loads prominently on the Level 1 Index also indicating the importance of configuration but also its association with other land use metrics. However, when disaggregated into the Level 2 index, it loses its significance in the statistical models. This brings into consideration not only that land use indices operate closely together in the landscape but also that they hold together more cohesively when combined rather than analyzed separately. In essence, the level 1 index might be a more robust measurement of landscape patterns.

Tract level land use variables remain significant after adding county-level random effects. However, it is interesting to note the direction of the coefficients associated with the county-level random intercepts. It appears that for certain counties, there are effects at the county scale that reduce/increase the risk of lung cancer mortality. There are an infinite number of county-level factors that may be associated with this county-level effect. The sample size limitations (only 21 counties) prohibit the addition of further fixed-effects at the county level. While beyond the scope of this dissertation, they are worth investigating in future research when data on more counties/tracts are available.

The models averaged a classification accuracy of around 70%. An analysis of Pearson residuals indicates that the residuals are not spatially autocorrelated. Majority of the misclassification occurs in the smaller tracts. Further analysis (addition of explanatory variables) is required to potentially improve classification accuracy.

The next section of the chapter present research findings in the literature that illuminate the relationships and mechanisms through which agricultural land uses poses a risk for lung cancer and situates it contextually within the study region and the state of Georgia.

Agriculture in Georgia

Agriculture is a major contributor to the Georgia economy, contributing $72.5 billion dollars as of 2012. The poultry and egg industry is significant and accounts for 47% of Georgia's farm commodities (Flatt, 2015; USDA, 2015). Georgia ranks first nationally in the production of broilers (young chickens). In Figure 39 we see that agricultural land uses are present at the peripheral tracts of the study area, almost in a circular ring. Similar coincidental patterns are seen in the spatial distribution of workers in the agriculture, forestry, fishing, hunting and mining occupations (Figure 40), spatial distribution of lung cancer mortality (Figure 41) as well as spatial distribution of COPD mortality (Figure 42). While the data does not permit a more fine-grained or causal investigation, it lends some evidence to the potential connection of agriculture and negative health outcomes, in this case lung cancer.

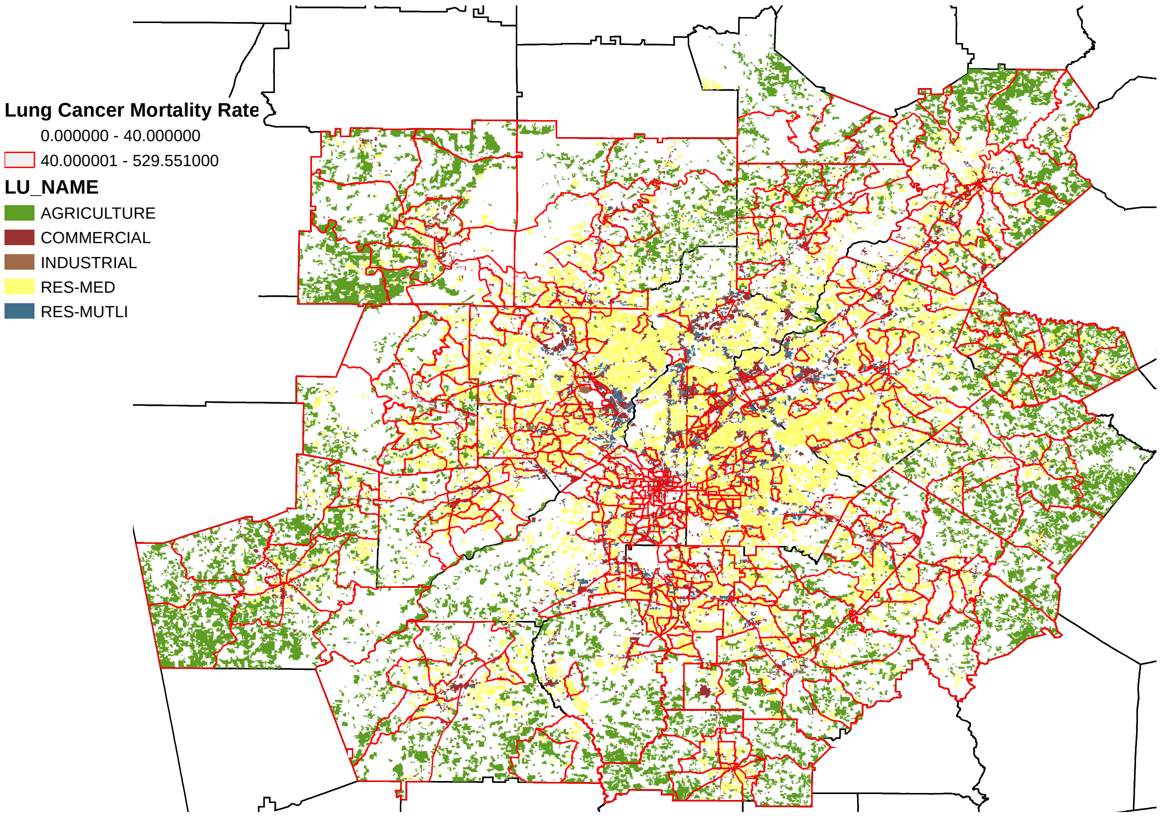


Figure 39. Spatial distribution of agricultural land use in the Atlanta Metropolitan area

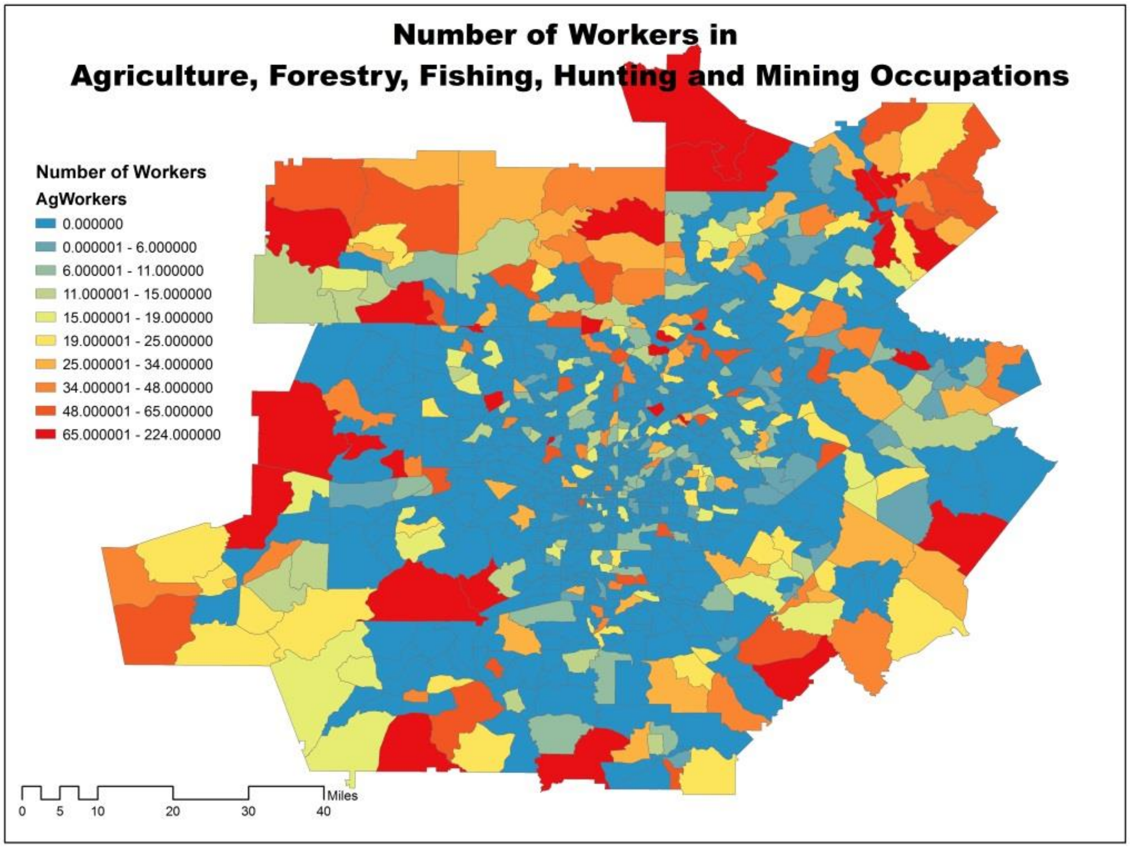


Figure 40. Spatial distribution of agricultural workers in the Atlanta Metro Area (Data source: American Community Survey, 2010)

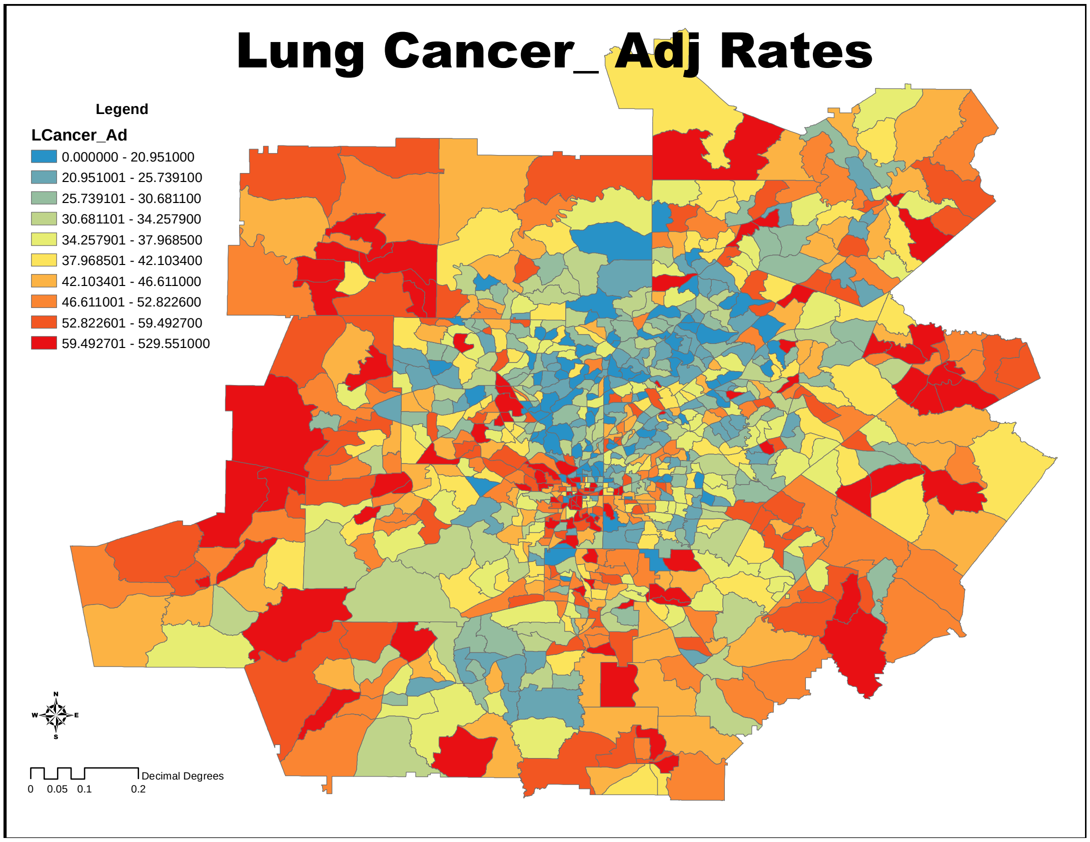


Figure 41. Spatial distribution of age-adjusted lung cancer mortality rates in the Atlanta Metro Area

Agriculture and Cancer- Land Use as exposure

The planning literature typically focuses on land uses as an enabler of healthy lifestyles (physical activity, access to healthy food) which in turn impact health outcomes such as obesity and other chronic diseases. However, these findings make an important contribution in potentially recognizing land use as an exposure that increases the risk for disease. It also reveals the potential for certain land uses and their patterns to serve as spatial signals for surveillance and monitoring programs that directly link environmental/occupational exposures to health outcomes. The primary mechanism associated with Agricultural land uses and lung cancer is through increased risk from pesticide exposure.

Pesticide exposure is recognized in the literature as an important risk factor in cancer development (Rull and Ritz, 2003; Jaga and Dharmani, 2005; Alavanja et al, 2007; Weichenthal et al, 2010). Common pathways through which pesticide exposure occurs include drinking water, air, dust and food. Exposure occurs directly for agricultural workers, particularly those who are involved in pesticide application. Ambient exposure occurs due to agricultural spray drift, posing an exposure risk to residents around agricultural land uses. Additional exposure occurs through occupational take-home pesticide residue on clothes and shoes.

All of these factors converge for residents of agricultural communities. Drift distances are greatly influenced by factors such as wind velocity, discharge height, ambient temperature, relative humidity and droplet size. Under certain conditions, pesticide droplets have been seen to drift upto 1000 ft.(Zhu et al, 1994). Several studies that utilize biological and environmental monitoring have found high levels of pesticides in samples of body fluids as well as indoor environments of individuals living in close proximity to agricultural areas (Ritz and Rull, 2008; Alavanja et al, 2007). For example, a study by Bradman et al (2005) showed that pregnant women living in agricultural areas had 2.5 times 2.5 times higher urinary metabolite levels of organophosphate pesticides compared to the general United States population.

Cancer risks for agricultural workers from pesticides

The dataset used for analysis in this dissertation reports on the final cause of mortality. In the case of lung cancer (and other cancers), the dataset does not specifically acknowledge if the cancer was metastatic or if the lung was the primary cancer site. While there is evidence in the literature linking agricultural exposures and increased risk of lung cancer, it is worthwhile briefly addressing the connection with all cancers.

The literature supports the linkage between increased risk of lung cancer in agricultural workers (Blair and Freeman, 2009; Weichenthal et al, 2010; McHugh, 2010). Specific pesticides implicated in this relationship include arsenic based compounds, organophosphate and carbamate insecticides as well as phenoxyacetic acid herbicides.

The Agricultural Health Study (AHS) is a prospective study of cancer and other health outcomes in a cohort of licensed pesticide applicators and their spouses from Iowa and North Carolina (http://aghealth.nih.gov/about/). Collectively, 89,000 farmers and their spouses enrolled in the study between 1993-1997. Follow-up studies regarding cancer and mortality have been conducted every five years since initial baseline at enrollment. Several scientific papers published using this dataset have shown elevated risks (upto two-fold) for several cancers including those of the lung (Alavanja et al, 2004; Weichenthal et al, 2010; McHugh et al, 2010; Jones et al, 2015; Lerro et al, 2015). The increased risk extends to spouses of agricultural workers as well (Lerro et al, 2015).

Chronic Obstructive Pulmonary Disease (COPD) and Landscape Patterns

COPD is considered a heterogeneous disorder or group of disorders, comprised of asthma, chronic bronchitis, emphysema, and airflow obstruction all being important parts of the final disease process (Mannino, 2002). In Chapter 4, the correlation analysis of health outcomes indicated a positive correlation between Lung Cancer and COPD. By itself, COPD accounted for approximately 5% of deaths between 2002 and 2011 in the study area. The risk factors for COPD closely resemble those for Lung Cancer. While smoking is the primary risk factor, environmental and occupational exposures play an important role in COPD causation (Raherison and Girodet, 2009; Mannino, 2002).

The socioeconomic gradient plays a significant role in COPD mortality and morbidity, similar to other chronic diseases. SES impacts COPD prevalence thorugh health behaviors and environments. Disparities in smoking prevalence, cessation, and disease burden have been persistently observed along the socioeconomic gradient. (Businelle et al, 2010; Hisock et al, 2012). Hypothesized mechanisms between low SES and high COPD include a spectrum of factors including prenatal exposures, more frequent lower respiratory tract illness in childhood, neighborhood disadvantage, housing conditions, air pollution, environmental tobacco smoke, diet, and other lifestyle factors including smoking, in addition to possible genetic factors (Prescotta and Vestbob, 1999; Businelle et al, 2010).

Ko and Hui (2012) summarize evidence from several studies that show increased risk for COPD prevalence and COPD exacerbation through exposure to outdoor pollution. Cumulative exposure over several years to PM10 and NO2 have been shown to increase risk for COPD prevalence. Studies have also shown more acute effects on COPD-related hospital admissions and emergency department visits (Ko and Hui, 2012; Schikowski et al, 2005; Arbex et al, 2009).

The relationship between COPD and agricultural exposures is similar to that of Lung Cancer. In a study on agricultural workers from 24 states across the US, the CDC (2007) reported significantly elevated mortality for several respiratory diseases, including tuberculosis, hypersensitivity pneumonitis, asthma, COPD, pneumonia, and influenza as measured by Proportionate Mortality Ratios (PMRs). Significantly elevated PMRs for COPD in particular were seen in crop and livestock farm workers. In addition to pesticides, repeated exposure to dust and gases on a regular basis contribute to this association (Kirkhorn and Garry, 2000; Schenker, 1998; Brackbill et al, 1994). Nordgren and Bailey (2016) note that 30% of COPD prevalence can be attributed to occupational exposures of which agricultural exposure form a significant source. They also note evidence in the literature that identifies socioeconomic status as a potential confounder in the link between farmers and COPD occurrence.

The literature also points to a close relationship between COPD and Lung Cancer. Raviv et al (2011) note that patients with COPD are at increased risk of developing Lung Cancer as well as poorer outcomes after Lung Cancer diagnosis and treatment. This is due to the fact that lung function is greatly impaired and has negative implications for treatment effectiveness and prognosis. There are many pathways that link COPD and lung cancer prevalence. Smoking, a risk-factor for both conditions, is one primary pathway. Chronic inflammation due to COPD is considered another primary mechanism that increases the risk for lung cancer, similar to chronic inflammation that triggers malignant transformations in other organs. This pathway is sometimes considered more dominant as it is prominently seen in non-smokers as well. COPD and lung cancer can occur simultaneously or COPD can precede it. The presence of other comorbidities such as emphysema and asthma confound this relationship as well (Durham and Adcock, 2015; Houghton, 2013; Barnes and Adcock, 2011; Koshiol et al, 2009).

The preceding paragraphs provide an overview of the various mechanisms and pathways for COPD prevalence and exacerbation. The brief literature review highlights environmental determinants that impact this relationship. It also highlights the fact that COPD and lung cancer share several risk factors, both can occur as comorbidities or COPD can be a precursor (risk factor) in itself to lung cancer prevalence. The next section provides an overview of the modeling results. Owing to its similarities to the lung cancer analysis and results, only key models and findings are explained.

Spatial Pattern of COPD mortality in the study area

Figure 42 shows the age-adjusted mortality rates for COPD mapped as deciles. A visual assessment of the rates indicates a peripheral ring of higher mortality rates. The adjusted mortality rates were then converted to the binary outcome variables, based on the methodology outlined in chapter 4. A side-by-side comparison (Figure 43) of lung cancer and COPD binary classification indicate a similarity in the pattern. Additionally, the literature review indicates shared risk factors and mechanisms. Hence the decision was made to include the results of the analysis as an addendum to lung cancer and briefly compare the outcomes of the modeling process.

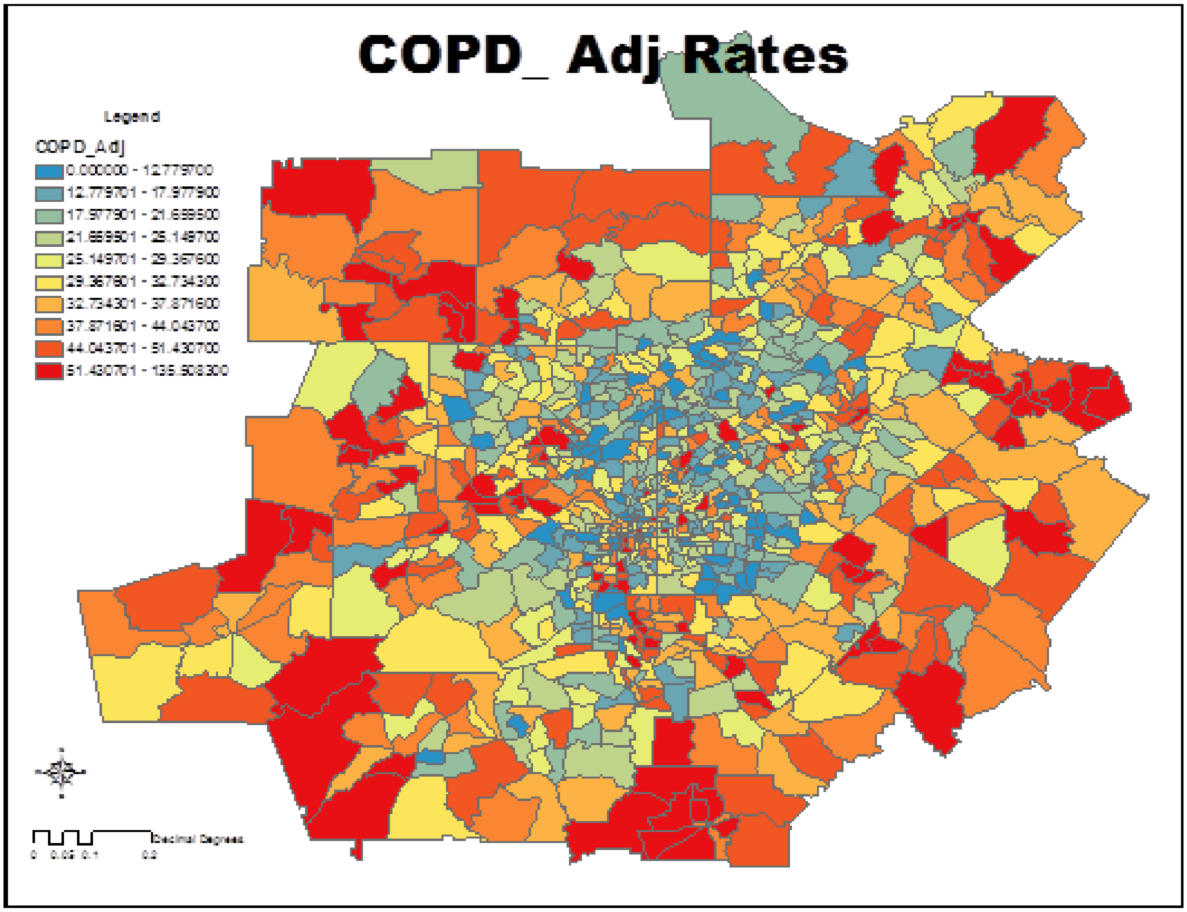


Figure 42. Spatial distribution of age-adjusted COPD mortality rates in the Atlanta Metro Area

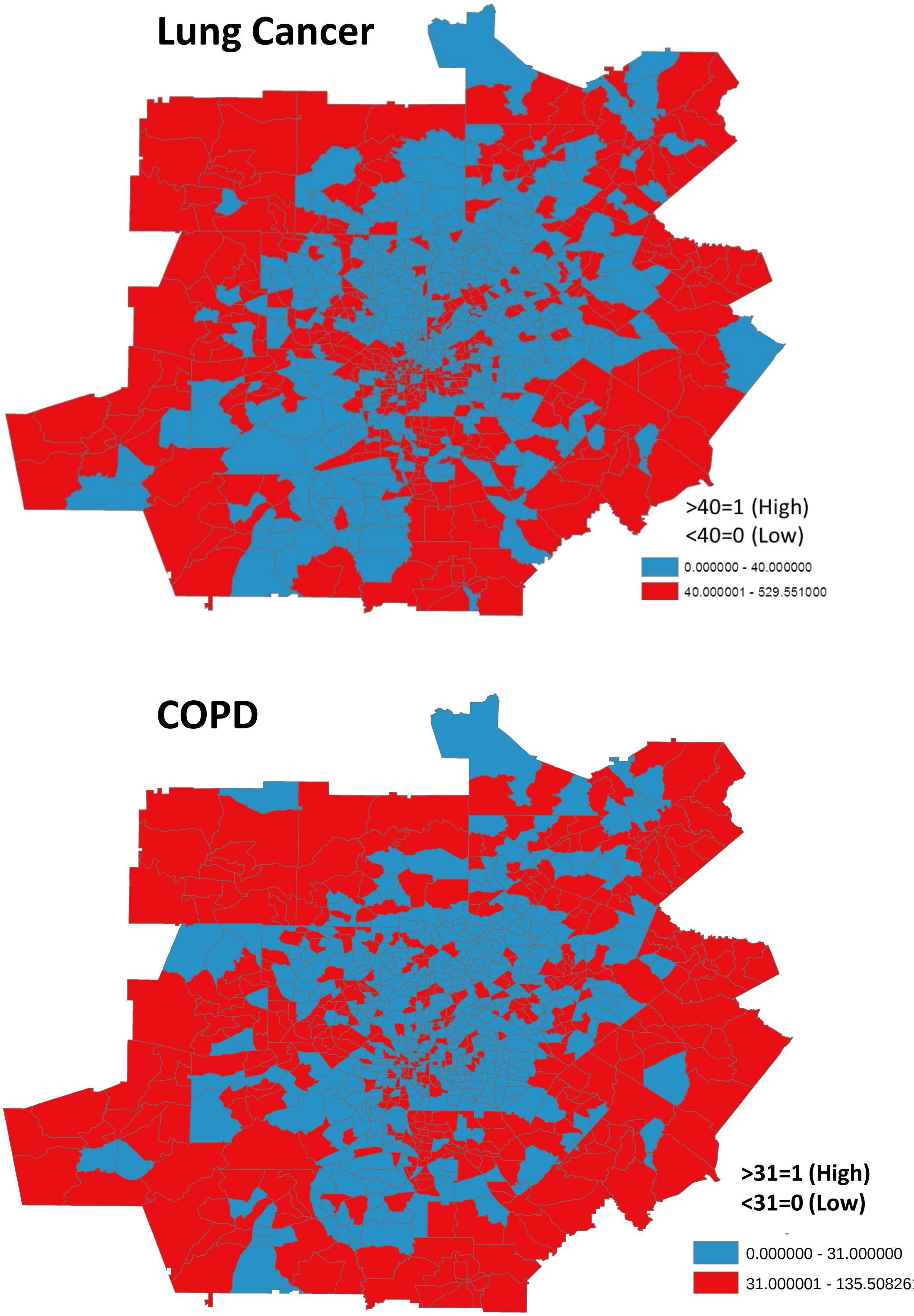


Figure 43. Comparative maps showing the spatial distribution of the binary outcome variables for Lung Cancer and COPD mortality

Exploratory and Confirmatory Modeling

Level 1 Variable Importance: Random Forest

Table 19. Level 1 Variable Importance: Random Forest

Level 1 Indices	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Random Forest Variable Importance	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini
SDI	9.786	26.552	11.133	24.534	6.701	24.378	6.549	24.876	9.637	25.179
SEI	7.338	23.624	9.627	23.004	7.861	23.694	8.698	24.599	6.575	22.239
SOCIAL	12.445	33.975	19.571	39.397	16.242	39.983	15.722	37.494	16.475	36.694
Index_AGRICULTURE	17.244	24.927	20.345	25.482	19.936	25.573	18.275	24.675	19.477	25.920
Index_COMMERCIAL	1.441	24.930	6.968	26.422	1.951	22.334	3.308	26.516	6.598	25.878
Index_FOREST	15.050	30.363	20.528	33.507	18.216	33.854	20.282	35.002	16.516	29.629
Index_IND_COM	8.151	16.279	3.933	12.261	4.026	12.897	3.793	12.098	7.189	13.775
Index_INDUSTRIAL	2.336	7.044	2.559	8.248	1.009	6.698	5.555	7.199	7.255	8.350
Index_INST_EXTENSIVE	0.150	7.858	-1.527	6.740	0.650	6.385	0.182	6.317	3.391	7.112
Index_INST_INTENSIVE	-2.651	19.960	0.189	20.161	-0.372	20.013	-1.585	18.658	1.310	19.579
Index_LTD_ACCESS	-1.021	10.063	4.187	12.532	0.463	11.788	0.851	11.004	3.931	13.301
Index_PARK_LANDS	-1.990	9.666	0.367	9.378	0.716	10.025	-1.041	10.067	0.230	9.363
Index_PARKS	1.126	14.738	0.701	15.452	1.721	14.800	1.188	14.594	3.247	14.799
Index_RES_LOW	11.475	24.695	11.672	22.053	11.564	23.247	11.431	24.311	15.821	26.918
Index_RESERVOIRS	2.006	15.311	2.997	14.572	1.833	13.766	6.245	15.021	4.919	13.247
Index_RES_HIGH	6.684	20.050	7.584	19.545	7.596	21.069	6.166	17.671	6.791	19.039
Index_RES_MED	0.604	22.327	2.772	20.747	2.216	22.160	0.993	22.381	4.320	21.563
Index_RES_MULTI	7.828	20.934	4.637	19.692	3.161	20.723	7.555	21.022	8.710	21.027
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	288	101	299	90	300	89	294	95	299	90
1	134	191	131	194	138	187	137	188	120	205
Testing, Actual **0**	103	26	92	37	99	30	95	34	98	31
1	42	66	52	56	39	69	45	63	48	60
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy [%]	**67.1%**	**71.3%**	**69.0%**	**62.4%**	**68.2%**	**70.9%**	**67.5%**	**66.7%**	**70.6%**	**66.7%**

Stepwise Logistic Regression with Level 1 Land Use Indices

Table 20.Results of Stepwise Logistic Regression with Level 1 Land Use Indices

STEPWISE LOGISTIC (fow/back	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	-1.7524	0.00038	-1.3584	0.01537	-0.19377	0.0207	-1.10264	0.041827	-0.21465	0.01255
SOCIAL	0.19973	0.00000	0.24191	0.00000	0.20459	0.00000	0.21217	0.00000	0.26451	0.00000
Index_AGRICULTURE	0.10829	0.01650	0.13846	0.00361	0.19893	0.00002	0.14523	0.00229	0.14232	0.00149
Index_COMMERCIAL	0.17258	0.00030	0.19188	0.00009	0.18283	0.00006	0.16300	0.00032	0.20315	0.00002
Index_RES_MULTI	-0.16799	0.00011	-0.12030	0.00436	-0.08781	0.03330	-0.13359	0.00169	-0.13630	0.00110
Index_RES_LOW									0.13882	0.00268
Index_LTD_ACCESS	-0.07618	0.01403	-0.07676	0.01510	-0.07071	0.02180	-0.06589	0.03395	-0.09253	0.00300
Index_IND_COM			0.04840	0.11279	0.06029	0.03720			0.07822	0.01232
Index_INDUSTRIAL	0.06303	0.03081	0.04563	0.13483			0.06049	0.04302	0.06416	0.04434
Index_RES_HIGH			-0.05912	0.08558					-0.05017	0.13348
Index_PARK_LANDS	-0.05118	0.10950							-0.04716	0.15323
SDI	1.06336	0.00144								
Index_FOREST	0.10810	0.02543	0.16066	0.00298	0.08520	0.09240	0.13448	0.00794		
SEI			1.75825	0.03849			1.40008	0.08603		
Index_INST_INTENSIVE			-0.07598	0.06552						
Index_RES_MED			0.08052	0.08239						
Confusion Matrix, Predicted	0	1	0	1	0	1	0	1	0	1
Training, Actual 0	287	102	292	97	292	97	290	99	302	87
1	123	202	122	203	130	195	127	198	118	207
Testing, Actual 0	100	29	93	36	93	36	95	34	88	41
1	47	61	54	54	47	61	42	66	48	60
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Classification Accuracy [%]	68.5%	67.9%	69.3%	62.0%	68.2%	65.0%	68.3%	67.9%	71.3%	62.4%

Multilevel Logistic Regression with Level 1 Land Use Indices

Table 21. Results of Multilevel Logistic Regression with Level 1 Land Use Indices

MULTILEVEL LOGISTIC	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	-1.14960	0.04906	-0.97644	0.11860	0.19252	0.41557	-0.71691	0.24625	0.05808	0.76988
SOCIAL	0.23088	0.00000	0.25658	0.00000	0.20933	0.00000	0.25990	0.00000	0.28517	0.00000
Index_AGRICULTURE	0.04978	0.34808	0.08678	0.12260	0.13266	0.02107	0.11643	0.03737	0.09494	0.07170
Index_COMMERCIAL	0.12788	0.01251	0.15957	0.00190	0.14897	0.00214	0.12663	0.00857	0.17131	0.00054
Index_RES_MULTI	-0.13525	0.00360	-0.10585	0.01720	-0.07389	0.08416	-0.11797	0.00890	-0.11841	0.00587
Index_RES_LOW									0.11554	0.02264
Index_LTD_ACCESS	-0.06357	0.05515	-0.05874	0.07920	-0.05927	0.06916	-0.04668	0.15565	-0.08609	0.00821
Index_IND_COM			0.03803	0.23170	0.06256	0.03669			0.07143	0.02685
Index_INDUSTRIAL	0.05853	0.05707	0.04577	0.14200			0.06078	0.05171	0.06439	0.04879
Index_RES_HIGH			-0.05801	0.14600					-0.04554	0.22271
Index_PARK_LANDS	-0.04078	0.24523							-0.04262	0.22531
SDI	0.95918	0.00758								
Index_FOREST	0.05793	0.26139	0.12813	0.02240	0.05880	0.26581	0.08224	0.11593		
SEI			1.69729	0.05430			1.35713	0.12107		
Index_INST_INTENSIVE			-0.04529	0.30160						
Index_RES_MED			0.04195	0.39540						
Random Effects - County	**Estimate**		**Estimate**		**Estimate**		**Estimate**		**Estimate**	
Barrow	0.43347		0.42802		0.32636		0.50467		0.15151	
Bartow	0.75760		0.61070		0.56051		0.27473		0.40671	
Carroll	0.34893		0.15368		0.33459		0.01179		-0.00324	
Cherokee	0.02958		0.04116		-0.22580		0.08832		0.09084	
Clayton	0.11876		0.23155		0.22469		0.25515		0.26630	
Cobb	0.02661		-0.08173		-0.12042		0.14707		-0.04593	
Coweta	0.27869		-0.03065		0.07023		0.26471		0.17358	
Dawson	-0.58368		-0.22684		0.05868		-0.25350		-0.13230	
DeKalb	-1.33614		-0.97542		-0.97950		-1.02149		-0.81973	
Douglas	-0.03802		0.02121		0.12614		0.25159		-0.02420	
Fayette	-1.41560		-0.77995		-1.05304		-0.92422		-0.87604	
Forsyth	-0.49837		-0.35141		-0.19560		-0.44447		0.05701	
Fulton	-1.17822		-0.98661		-0.88536		-1.11207		-0.84760	
Gwinnett	-0.47476		-0.21465		-0.54129		-0.04424		-0.18165	
Hall	0.27636		0.05208		0.12497		-0.26986		0.06232	
Henry	0.75023		0.34873		0.18287		0.16455		0.20169	
Newton	0.68795		0.16847		0.67110		0.51033		0.29067	
Paulding	-0.03555		0.04676		0.08281		0.15223		0.19840	
Rockdale	0.01230		0.29321		0.03784		0.08099		0.05769	
Spalding	0.73063		0.45197		0.52718		0.48698		0.39298	
Walton	0.67894		0.57228		0.41140		0.61167		0.42699	
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	348	41	344	45	353	36	349	40	348	41
1	154	171	166	159	173	152	163	162	161	164
Testing, Actual **0**	112	17	115	14	116	13	107	22	116	13
1	58	50	65	43	65	43	57	51	60	48
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy	**72.7%**	**68.4%**	**70.4%**	**66.7%**	**70.7%**	**67.1%**	**71.6%**	**66.7%**	**71.7%**	**69.2%**

Level 2 Variable Importance: Random Forest

Table 22. Level 2 Variable Importance: Random Forest

Level 2 Indices	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Random Forest Variable Importance	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini	Mean Decrease Accuracy	Mean Decrease Gini
SOCIAL	16.464	49.659	17.813	52.558	19.164	52.084	15.976	49.240	27.536	55.669
IDX_GEOM_AGRICULTURE	18.777	29.709	14.780	24.492	16.087	26.671	12.277	26.909	13.171	25.921
IDX_SHAPE_AGRICULTURE	15.409	25.229	14.055	24.306	14.864	26.411	16.647	27.988	17.888	28.366
IDX_INTER_AGRICULTURE	12.364	23.318	14.736	24.950	16.632	24.948	12.236	24.064	14.066	24.615
IDX_GEOM_COMMERCIAL	7.678	34.043	9.591	36.800	8.750	35.169	3.149	32.970	7.819	35.418
IDX_SHAPE_COMMERCIAL	-2.048	30.096	-0.004	30.370	0.481	29.494	0.489	28.506	-1.798	27.470
IDX_INTER_COMMERCIAL	4.177	29.069	1.547	30.503	3.111	30.546	3.983	30.743	2.243	29.295
IDX_GEOM_INDUSTRIAL	7.528	7.893	7.666	6.464	7.441	7.776	4.304	7.734	8.630	7.222
IDX_SHAPE_INDUSTRIAL	4.113	7.272	4.113	6.148	2.041	5.571	1.676	6.631	4.299	5.136
IDX_INTER_INDUSTRIAL	6.849	6.887	7.803	5.821	8.451	6.838	7.877	6.401	8.263	5.047
IDX_GEOM_LTD_ACCESS	2.510	11.627	3.605	12.840	2.400	11.713	1.646	11.761	5.744	13.720
IDX_SHAPE_LTD_ACCESS	2.841	11.813	2.522	10.790	0.911	11.968	2.539	11.208	7.568	12.527
IDX_INTER_LTD_ACCESS	4.229	12.338	1.250	11.720	1.624	11.375	1.533	12.099	4.248	11.671
IDX_GEOM_RES_MULTI	4.859	23.358	2.871	24.293	2.612	23.222	6.006	24.777	5.086	23.150
IDX_SHAPE_RES_MULTI	2.952	23.821	1.964	24.539	2.794	25.124	8.944	27.639	2.495	23.587
IDX_INTER_RES_MULTI	3.744	24.118	3.535	23.860	0.760	22.193	0.961	22.830	4.789	21.897
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	290	99	286	103	287	102	291	98	289	100
1	140	185	134	191	131	194	148	177	133	192
Testing, Actual **0**	100	29	90	39	102	27	90	39	86	43
1	44	64	39	69	50	58	45	63	46	62
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy [%]	**66.5%**	**69.2%**	**66.8%**	**67.1%**	**67.4%**	**67.5%**	**65.5%**	**64.6%**	**67.4%**	**62.4%**

Stepwise Logistic Regression with Level 2 Land Use Indices

Table 23. Results of Stepwise Logistic Regression with Level 2 Land Use Indices

STEPWISE LOGISTIC (fow/ba	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	-0.17263	0.03941	-0.14968	0.07023	-0.19736	0.01938	-0.17470	0.03765	-0.17497	0.03809
IDX_GEOM_AGRICULTURE	0.28904	0.00000	0.27713	0.00000	0.27538	0.00000	0.27803	0.00004	0.27560	0.00004
SOCIAL	0.18850	0.00000	0.18320	0.00000	0.20732	0.00000	0.19447	0.00000	0.19686	0.00000
IDX_INTER_INDUSTRIAL	0.31503	0.00074	0.25434	0.00544	0.32151	0.00080			0.23746	0.01054
IDX_INTER_AGRICULTURE	0.22721	0.01294	0.26249	0.00340	0.22535	0.01287				
IDX_GEOM_COMMERCIAL	0.14520	0.00903	0.12701	0.01206	0.16435	0.00437	0.18911	0.00126		
IDX_SHAPE_RES_MULTI	-0.08374	0.08665								
IDX_SHAPE_LTD_ACCESS	-0.06614	0.09285					-0.06522	0.10186		
IDX_GEOM_RES_MULTI					-0.11761	0.04010	-0.09233	0.09384		
IDX_SHAPE_COMMERCIAL					0.11461	0.06973			0.10835	0.05167
IDX_SHAPE_AGRICULTURE							0.13974	0.01856	0.14871	0.00923
IDX_GEOM_INDUSTRIAL							0.13176	0.00368		
IDX_INTER_COMMERCIAL									0.22051	0.00679
IDX_INTER_RES_MULTI									-0.15985	0.06734
Confusion Matrix, Predicted	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
Training, Actual **0**	300	89	301	88	290	99	305	84	304	85
1	138	187	148	177	123	202	146	179	141	184
Testing, Actual **0**	97	32	97	32	97	32	93	36	95	34
1	48	60	40	68	53	55	46	62	50	58
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy [%]	**68.2%**	**66.2%**	**66.9%**	**69.6%**	**68.9%**	**64.1%**	**67.8%**	**65.4%**	**68.3%**	**64.6%**

Multilevel Logistic Regression with Level 2 Land Use Indices

Table 24. Results of Multilevel Logistic Regression with Level 2 Land Use Indices

MULTILEVEL LOGISTIC	Iteration # 1		Iteration # 2		Iteration # 3		Iteration # 4		Iteration # 5	
Variable Significance	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
(Intercept)	0.40097	0.13311	0.50752	0.06890	0.32012	0.20710	0.38333	0.15166	0.40910	0.15823
IDX_GEOM_AGRICULTURE	0.14058	0.05840	0.13792	0.04330	0.15482	0.03020	0.19371	0.01323	0.18375	0.02027
SOCIAL	0.20232	0.00000	0.22188	0.00000	0.22704	0.00000	0.21975	0.00000	0.22436	0.00000
IDX_INTER_INDUSTRIAL	0.30193	0.00204	0.21651	0.02390	0.20005	0.03160			0.25731	0.00988
IDX_INTER_AGRICULTURE	0.09030	0.38141	0.10766	0.27060	0.21616	0.04040				
IDX_GEOM_COMMERCIAL	0.11105	0.06433	0.11727	0.03100	0.14420	0.01700	0.15486	0.01209		
IDX_SHAPE_RES_MULTI	-0.05155	0.31841								
IDX_SHAPE_LTD_ACCESS	-0.06913	0.10564					-0.04484	0.29450		
IDX_GEOM_RES_MULTI					-0.08920	0.13500	-0.07869	0.16653		
IDX_SHAPE_COMMERCIAL					0.12416	0.05230			0.10071	0.08579
IDX_SHAPE_AGRICULTURE							0.02486	0.71123	0.04200	0.52016
IDX_GEOM_INDUSTRIAL							0.13522	0.00402		
IDX_INTER_COMMERCIAL									0.20419	0.02087
IDX_INTER_RES_MULTI									-0.16819	0.07872
Random Effects - County	Estimate		Estimate		Estimate		Estimate		Estimate	
Barrow	0.53135		0.24055		0.69438		0.64797		0.61883	
Bartow	0.88188		0.61731		0.65582		0.46496		0.61060	
Carroll	0.62649		0.50420		0.18068		0.96600		0.69585	
Cherokee	-0.21825		0.03969		0.11695		-0.23249		-0.04066	
Clayton	0.25053		0.05635		-0.13325		-0.00984		0.40503	
Cobb	-0.52470		-0.32374		-0.46306		-0.33571		-0.32060	
Coweta	0.58865		0.69098		0.06977		-0.05148		0.26385	
Dawson	-0.77377		-0.39799		-0.71562		-1.08179		-0.80537	
DeKalb	-1.40215		-1.63327		-1.24753		-1.52913		-1.32453	
Douglas	-0.37951		-0.18802		-0.15205		0.03737		-0.35874	
Fayette	-0.97276		-1.49059		-1.05349		-1.16758		-1.99574	
Forsyth	-0.37702		-0.51710		-0.49762		-0.35875		-0.49757	
Fulton	-1.39612		-1.52620		-1.19867		-1.37522		-1.63916	
Gwinnett	-0.64535		-0.67795		-0.40733		-0.40772		-0.39528	
Hall	0.34902		0.03122		0.34471		0.11464		-0.31235	
Henry	0.35332		0.49162		0.92959		0.48163		0.54716	
Newton	0.97707		0.76840		0.88601		0.44330		1.23643	
Paulding	-0.03552		0.81160		-0.01761		0.75857		0.03087	
Rockdale	0.24287		-0.03737		-0.01241		0.49909		0.58034	
Spalding	0.62222		0.88102		0.77735		0.70109		0.99180	
Walton	0.78815		1.02144		0.78284		0.86688		0.99467	
Confusion Matrix, Predicted	0	1	0	1	0	1	0	1	0	1
Training, Actual 0	353	36	349	40	356	33	353	36	354	35
1	165	160	171	154	170	155	170	155	157	168
Testing, Actual 0	117	12	106	23	115	14	109	20	112	17
1	68	40	52	56	60	48	59	49	67	41
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Classification Accuracy [%]	**71.8%**	**66.2%**	**70.4%**	**68.4%**	**71.6%**	**68.8%**	**71.1%**	**66.7%**	**73.1%**	**64.6%**

Analytical Summary

In the Stepwise logistic model, The Neighborhood Deprivation Index (Social) is highly significant and has the highest co-efficient. The level 1 Index for Agriculture is also significant though the level of significance varies between iterations. Other significant variables are Commercial, Multifamily Residential and Limited Access (highways). The negative association for Multifamily Residential and Limited Access needs to be examined further. One potential explanation is that at the tract scale, there might not be residential areas around highways (low populations and hence low rates). In the multilevel model, the Neighborhood Deprivation Index (Social), Commercial and Multifamily Residential stay consistently significant. The level 1 indices for Agriculture and Limited Access are less consistent across iterations, potentially indicating the impact of county-level effects.

The level 2 analysis yielded interesting results. According to the Random Forest and Stepwise Logistic Models, the Agriculture Geometry Index is even more significant compared to the Neighborhood Deprivation Index (Social). The Interspersion indices for Agriculture, Industrial and the Geometry Index for Commercial are occasionally significant among iterations. The Neighborhood Deprivation Index (Social) regains its importance as the most significant variable in the multilevel model. Next, the Geometry Index for Agriculture stays consistently significant. Again, the Interspersion indices for Agriculture, Industrial and the Geometry Index for Commercial are occasionally significant among iterations.

The county-level effects remain very similar to those seen in the Lung Cancer models. Cherokee, Cobb, Dawson, DeKalb, Douglas, Fayette, Forsyth, Fulton, Gwinnett and Paulding counties have negative intercepts, indicating an ameliorating effect on tract-level health outcomes.

CHAPTER VI

IMPLICATIONS OF FINDINGS FOR URBAN PLANNING AND HEALTH RESEARCH

Introduction

The preceding chapters laid out the rationale, conceptual framework, a detailed methodological approach and modeling results for two disease outcomes as illustrative applications for this approach. Chapters 1 summarized the motivation and objectives for this research project. Chapter 2 provided an overview of the key theoretical bodies of work that contributed to the conceptual and methodological framework. Chapter 3 summarized the entire analytical approach, outlining the progression from raw data to confirmatory analysis. Chapter 4 dove into details of the actual datasets, their conversion into the dependent and independent variables as well as a descriptive investigation of their spatial patterns. Chapter 5 provided a demonstration of this entire framework to the disease outcomes of lung cancer and COPD mortality rates.

The entire framework developed in this document is derived for the Atlanta region (21-county area as defined by the Atlanta Regional Commission). The details provided regarding the methodology and approach make it easily replicable in other geographic areas, based on data availability. It is also applicable to other health outcomes and urban planning objectives (such as transportation planning). Chapter 6 provides a roadmap of how this framework can be replicated and applied to health and other urban issues. It also provides recommendations on how the methods and findings can be consumed to extend research on healthy places.

Summary of Findings

The research undertaking explained in this dissertation can be broadly categorized into three steps—data acquisition, data processing and data analysis. Accordingly, three levels of findings from this dissertation are relevant for discussion, as they provide valuable insights for planners who might wish to venture on similar endeavors:

Findings relevant to acquisition and consumption of data

Evidence-based policy making is heavily reliant on sound data and methods. Data and statistical modeling form a critical foundation to the framework presented in this dissertation. A discussion of policy implications really begins with the underlying data, it's limitations, modeling assumptions and what is included/omitted in the process.

Obtaining and consuming census tract level health data

Publicly available health datasets are great assets for research. National datasets such as the County Health Rankings provide great opportunities to detect trends in counties over time and across the nation. Other similar datasets exist with various levels of permissions for access. Counties are usually the smallest geographic unit at which health data is publicly released. While this enables robustness for research and better comparability across units, the spatial heterogeneity within counties is not captured.

Planners tend to favor using the finest scale of data available. Census tracts (block groups and even blocks) offer a better glimpse at intra-county variations in health outcomes. However, the utilization and consumption of data at these finer scales requires caution. Annual mortality and morbidity indicators may sometimes represent anomalies rather than meaningful temporal patterns.

Additionally, rates calculated from these counts are subject to the “small number” problem. Rates are calculated based on denominators generally composed of the population counts of interest. Denominators (population totals) tend to get smaller with smaller geographic units. Rates calculated with these smaller denominators, even if they have small numerators, provide inflated measures of risk when none might exist. Models built on these short timespans may be subject to spurious associations. Methods commonly used to deal with these issues include data suppression, smoothing or aggregating counts over larger time periods.

In the state of Georgia, the Georgia Department of Public Health (GaDPH) collects, stores and shares data. The data is made accessible through a formal data request process which requires justifications for the data and the scale at which it is requested. This aligns with data and human subjects protections as required by HIPAA and IRBs. The GaDPH also maintains the Online Analytical Statistical Information System (OASIS) which provides data on several health indicators. While rates and adjusted rates are publicly available at the county and larger scales, census tract data are only available as descriptive categorical maps. Furthermore, rate calculations cannot be performed on this type of data.

The data and methods described in this dissertation provide a potential model of processing, analyzing and consuming finer scale data. The first step in obtaining data was to submit a detailed data request. The data request required a detailed explanation of the research questions, methods and IRB clearance. Planning agencies find it challenging to obtain detailed health data. In the case of Georgia, the data exists but barriers to accessibility also exist. However, the complex nature of health data and scale pose mathematical challenges that require technical gatekeepers.

The current policy environment encourages a much broader conception of environment and health relationships based on the socioecological model. For example, the Accountable Care Act poses several opportunities for collecting community level health data through the Community Health Needs Assessment process. The challenge for regional and local planning agencies is coordinating and triangulating the efforts and its findings and incorporating it within planning processes. However, utilizing this information provides prime opportunities for framing focused research questions to obtain detailed data and creating evidence-based interventions. **Thus, the first step in obtaining health data at finer scales is prioritizing health needs and framing it as a formal research effort rather than a database expansion project**. This helps planning agencies be successful with data requests and building a foundation for potential long-term data sharing agreements.

The research presented here deals with the mathematical challenges of census tract level health data outlined in the following paragraphs. First, the data was provided as 10 year aggregate mortality counts for a series of diseases. Aggregation over 10 years is one technique of smoothing over unpredictability of rates at smaller geographic levels.

Second, mortality measures have certain advantages over morbidity rates. With current healthcare and technology, there are undoubtedly many steps between getting sick and dying. As urban planners, we strive to improve quality of life rather just prevent deaths. However, morbidity data is also harder to access and is prone to sampling biases. In Georgia, it requires additional permissions from the Georgia Hospital Association.

Mortality data might be less susceptible to reporting/sampling bias and can be considered more robust for associative and predictive modeling.

Finally, age-adjusted mortality rates were calculated and utilized for modeling purposes. Age is by far the most significant factor that is correlated with mortality. Thus, age-adjusting is a critical step in ensuring that the significance of all other explanatory variables (such as built environment) is valid and not merely a function of the underlying demographic distribution. Age-adjusting also makes tracts truly comparable with respect to their environmental variables as the variation due to age has been factored into the dependent variable. The calculated age-adjusted rates were then compared with the county level age-adjusted rates to ensure that there was consistency between spatial patterns seen at these two scales.

Advantages and disadvantages of the land use data

Detailed descriptions of the LandPRO 2010 dataset are provided in previous chapters. The dataset is an on-screen photo-interpretation and digitizing of ortho-rectified aerial photography at a scale of 1:14,000. The primary source for this GIS database was 2009 true color imagery with 1.64-foot pixel resolution. Combined with supplemental ownership information, it provides a generalized vector dataset of 25 land use types. 15 land uses out of the25 were used for further analysis after a numerical and theoretical assessment of their significance to the study.

The accuracy of the generalization directly impacts the validity of landscape metrics at the tract level. ARC uses a 5 acre mapping unit for differentiation between land uses. The smallest tracts within the study area have an approximate area of about 375 acres. This still permits capturing sufficient variation in land uses based on the unit of generalization used.

The Atlanta Regional Commission (ARC) publishes this dataset for public use on a regular basis (latest being 2010). Since 1999, several process improvements have been incorporated to improve the positional accuracy of the data. The ARC recommends that this data can be used for regional and municipal transportation, environmental and landuse planning. However, it is not to be used outside of its regional context.

The LandPro data allows comparisons between counties within the Atlanta region. However, it is challenging to find and reconcile similar land use data across several regions outside of Atlanta. From a methodological perspective, the study could have been strengthened by using land use data from other parts of the country. However, several feasibility considerations prohibited this. Thus, the results generated from the application of this framework are applicable to the Atlanta Region only

While the results might be place-specific, the framework itself can be expanded to other regions. Unlike mortality classifications, there is much larger variation in land use classification categories and the methods with which they are constructed. Thus, the standardization of land use categories would be key to an expanding the geographical scope. One among many possible ways would be to standardize the interpretation/classification of satellite/aerial imagery across other regions under consideration. This way, the researcher has complete control over the accuracy and thresholds required to construct comparable datasets.

Advantages and disadvantages of the demographic and socioeconomic data

Census data was used to obtain demographic and socioeconomic variables of interest. The utilization of demographic data is an important point of discussion for this dissertation. The original health dataset contained mortality counts (numerator) for a ten year time period from 2002 to 2011. In order to convert this to a rate, a corresponding ten year population count had to be assembled for the denominator. This involved putting together decennial data (2000, 2010), American Community Survey estimates and interpolating between these for time points that did not have ACS estimates.

The mortality data was provided based on 2010 census tracts. However, a small number of tract geographies changed between the years 2000 and 2010. The Longitudinal Tract Data Base (LTDB) published by Brown University was used to download population estimates for the year 2000 with tract equivalencies based on 2010 tract geography. This is one area in the analysis where some estimation (and associated uncertainty) is introduced.

Age-adjusting further involved reconciling age groups between the GaDPH data and the census data. The age distribution for the Atlanta Metro area was used as the standard population to perform the age adjustment. This again makes the results of the analysis strictly applicable to the study area only. The tract rates were compared to county-level rates from Ga DPH to ensure validity.

Socioeconomic data from 2010 ACS estimates were used to construct the Neighborhood Deprivation Index. The variables were utilized in the form of percentages so that the variables and the resulting index would be comparable across tracts. Some variables with a lot of missing data were excluded from the analysis. The variables were used to construct a Neighborhood Deprivation Index, an area level measure of socioeconomic deprivation, for each tract. Principal Components Analysis was used as a way of dealing with the high degree of correlation between the variables.

Findings relevant to exploratory analysis (this includes construction of variables/ indices from the raw data as well as exploratory patterns, hierarchical clustering)

The age-adjusted mortality rates were mapped to visually assess spatial patterns. The age-adjusted rates were then converted to binary outcome variables based on the findings from the LISA cluster analysis (details in chapter 4). The conversion to binary variables offers two advantages. First, it sets up a contrast between high rates and low rates in the area. Differences and the significance of the differences in rates may not be apparent when the outcome is treated as a continuous variable. This can also be problematic when the range of mortality rates is numerically small. The threshold was derived from the LISA analysis and not arbitrarily assigned. Second, bias in the dataset is reduced when there are outliers present that could skew the results of the modeling process (2 tracts had higher than expected lung cancer rates).

Lung cancer and COPD had similar patterns with a ring of higher rates around the periphery of the study area. Rates for Heart Disease and Diabetes were similar, with a more diffused pattern. The initial study of correlations between disease rates coincides with these correlations between spatial patterns as well.

Landscape Metrics were computed through Patch Analyst using the ArcGIS platform. The metrics were consolidated into two levels of indices for each land use of interest as a way of reducing correlations between the metrics. Principal Components Analysis was used to construct Level 1 indices for each land use. Consequently each land use type had one index value for each tract. Level 2 indices were disaggregated versions for each of the land uses. Here, the metrics were grouped based on the aspect of the landscape they represented— namely, geometry, shape and interspersion. Geometry and shape can simultaneously represent composition and configuration whereas interspersion is a cleaner measure of configuration.

Exploratory analysis was conducted using Hierarchical Clustering. The full suite of Level 1 land use indices, Neighborhood Deprivation Index and lung cancer rates were included (Figure 44). At first glance, the pattern of signatures between the variables is not evident (too many unique signatures for a discernible pattern). While the relationship of health outcomes to the other variables is not as clean, the pattern between land use and neighborhood deprivation is more interesting. The independent variables were clustered in the diagram.

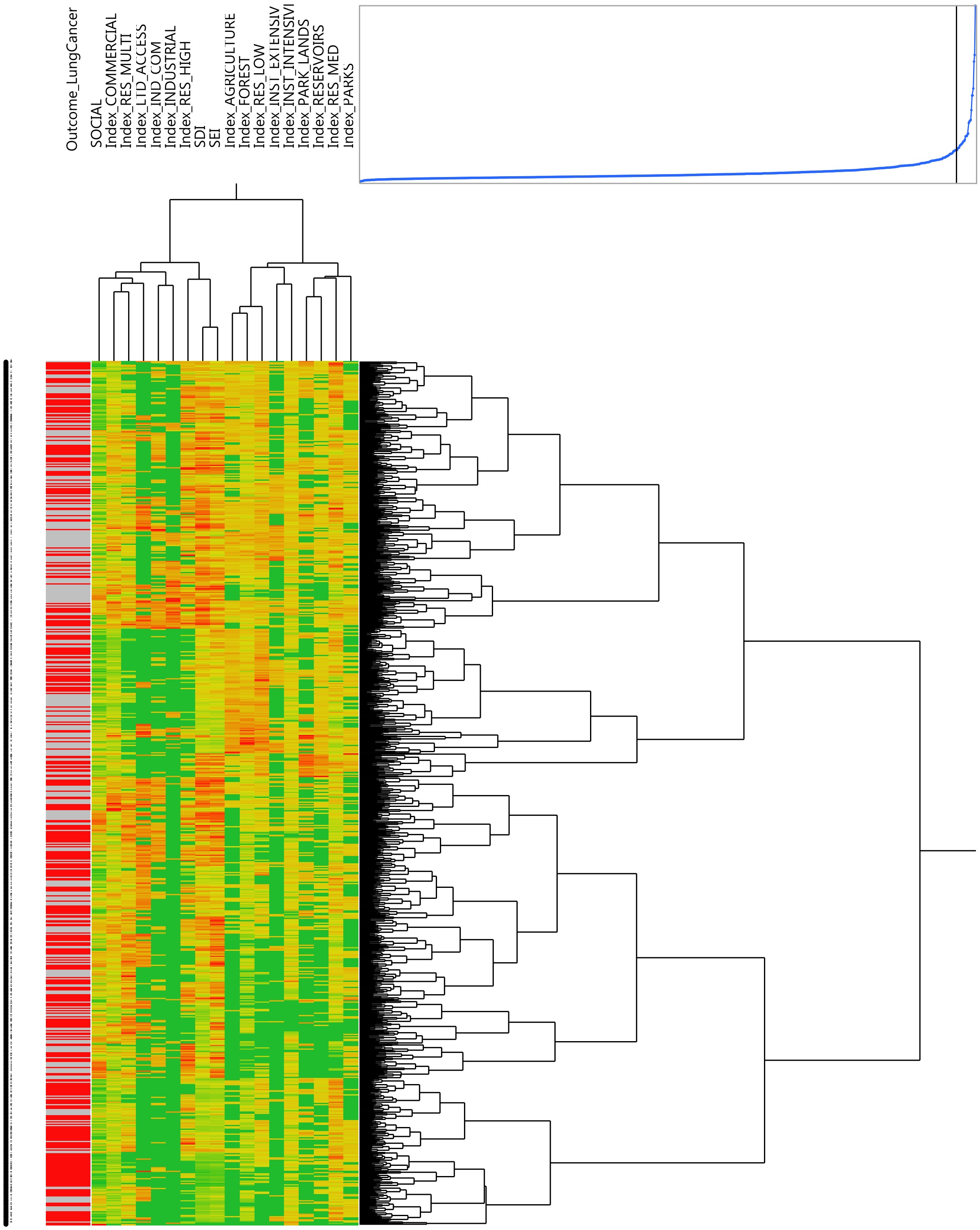


Figure 44. Hierarchical Clustering of lung cancer rates, Neighborhood Deprivation Index (Social) and Level 1 Land Use Indices

The clustering reveals that Neighborhood Deprivation has a closer relationship to Commercial, Multifamily Residential, Limited Access, Industrial/Commercial, Industrial and High Density Residential land uses. On the other hand, Agriculture, Forest and Low Density Residential clump together along with Parks and other land uses that represent conservation areas.

The most significant takeaways from this section are that two types of correlations exist. The first type include the **significant correlations amongst landscape metrics (of the same land use type) as well as socioeconomic variables**. The second kind of correlation pertains to certain types of land uses being more closely related to socioeconomic conditions compared to others. Again, it is important to remember that these results are applicable only to the study area and should not be taken out of context. **Principal Components Analysis** is an effective way of reducing the data into uncorrelated components. The first component almost explains a majority of the variance and **is used to construct the relevant index**.

One of the objectives of this dissertation was to look at **landscape signatures** or groupings of land use and socioeconomics that may help classify a place as healthy or unhealthy. The Hierarchical Clustering diagram was an effective exploratory tool that helped visualize these relationships. The **clustering patterns** indicate certain types of **socioeconomic and land use juxtapositions** that occur in the study area.

Findings relevant to the confirmatory modeling process

Age-adjusted rates were calculated and their patterns mapped for Lung Cancer, COPD, Heart Disease and Diabetes. However, Lung Cancer and COPD were examined in greater depth for the confirmatory modeling process. A rigorous approach was used to ensure consistency between models of the same type as well as across model types.

Each model used a randomly selected sample of training and testing data for every iteration. Six different types of modeling techniques were utilized to answer the research questions. This was also a good test of consistency between model types. Five iterations were run for each model type to check for agreement within models of the same type. The first three model types were run using the Level 1 aggregated land use indices. Variables that stayed consistently significant across model iterations and between model types were carried into the Level 2 analysis. Disaggregated sub-indices for each of the significant land use variables from Level 1 were used in the Level 2 analysis. All three model types were rerun with the sub-indices.

The Neighborhood Deprivation Index emerged as the most significant variable in all the models. However, the Agriculture index was the next most significant variable for both disease outcomes. Two types of secondary evidence were used to validate this finding. The literature documents several mechanisms through which occupational exposures in Agriculture pose a risk for Lung Cancer and lung diseases in general. Census data also confirmed that tracts with high rates also had a significant number of agricultural workers living there in addition to presence of agricultural land use. This aligned with the fact that mortality data was provided by place of residence.

A few other land use variables were also significant though their estimates were smaller in magnitude and their p-values were not as significant as lung cancer (all were still significant at 0.05 level). Based on the hierarchical clustering diagram, several of these land uses appear to occur in areas with high Neighborhood Deprivation. While this type of analysis is not confirmatory, it provides some insight for further analysis in future research.

There is statistical evidence from multiple models that land use variables have an impact on lung cancer risk, even after accounting for socioeconomic variables, adjusting for population age distributions and including county-level effects.

Interpreting the analytical results as answers to the Research Questions

Age-adjusted mortality rates and corresponding binary outcomes were calculated for four diseases (Lung Cancer, COPD, Heart Disease and Diabetes) that contributed prominently to the total mortality in the study area. Lung Cancer was examined in detail through the exploratory and confirmatory modeling process. Select models for COPD were also presented due its correlation with Lung Cancer Risk and similarity in modeling results.

The exploratory and confirmatory analysis sought to answer three principal research questions central to this dissertation.

1. Are landscape patterns important determinants of human health?

The importance (statistical significance) of land uses and their contribution to health outcomes were evaluated using stringent criteria. A land use co-efficient was determined to be significant ($p<0.05$) only if it passed all the following checks:

- Was the co-efficient consistently significant for every iteration in a model type?
- Was the co-efficient consistently significant across iterations and across model-types?

- How did the value of the estimate and p-value compare with the estimate for the Neighborhood Deprivation Index?

The Neighborhood Deprivation Index had by far the largest and most statistically significant association with disease risk compared to all variables. For the Lung Cancer models, the Agriculture index was in agreement with all of the above requirements. While a few other variables were significant, their co-efficient estimates and p-values, while still less than .05, were not as strong as Agriculture ($p<0.001$ in most models). Similar results are observed for COPD mortality as well.

Models using Level 1 land use indices were also analyzed for Heart Disease and Diabetes (not included in the dissertation document). As per the criteria outline above, only the Neighborhood Deprivation Index showed a strong and consistent association with mortality rates.

Based on the methodology and results presented above, we can conclude that Land Use has a differential impact based on the health outcome of interest. The socioecological model of public health purports that all diseases have an environmental component that influence their prevalence and associated mortality. However, the nature of that environmental envelope requires the conceptualization of models built on relevant risk factors.

In the case of Lung Cancer and COPD, the literature documents environmental/occupational exposures that influence disease outcomes. These exposures operate both independently and through interaction with socioeconomic factors. This is one potential reason that the Agricultural index is significant. Tracts with high lung cancer mortality also have high proportions of agricultural land use and residents employed in the agricultural sector, compared to other tracts in the study, area, strengthening this hypothesis.

In the case of Heart Disease and Diabetes, exposures accrued through deprived social environments play a more significant role. It is important to examine in greater depth the implications of the neighborhood deprivation index and what it represents. There is a vast amount of literature that documents neighborhood deprivation as a measure of both social and physical processes. Deprived neighborhoods are known to have poor food environments and reduced access to other resources that support healthy behaviors and better healthcare.

The discussion then really becomes, *"Are measures of land use and deprivation completely different or do they measure different facets of the same problem?"* While coarser measures of land use patterns (such as landscape metrics) may measure certain aspects of environmental exposures, finer measures of land use quality (such as measures that further distinguish commercial into grocery stores, healthcare, etc.) may be better captured by measures of deprivation.

The analysis also reveals certain types of land uses that are more likely to co-occur with higher deprivation. There is a weak but positive correlation between high deprivation and land uses such as Commercial, Industrial and Multifamily Residential. The hierarchical clustering diagram also shows evidence of this. While the relationship may be highly place-specific or non-linear, it is worth considering when creating signatures of healthy/unhealthy places.

Hierarchical Clustering can also be used to explore heterogeneity of landscape signatures within healthy and unhealthy tracts. Figure 45 shows different manifestations of landscape signatures for tracts that have high lung cancer mortality rates. 7 optimal clustering patterns are identified (as indicated by the bend in the scree plot). Distinct landscape signatures are identified based on the color of the cluster (for example, cluster 2 is identified by the green branches of the dendrogram). ID numbers are also assigned on the left.

The dendrogram can be interpreted as a matrix of the independent variables plotted against tract ids. Each cluster differentiates itself based on the unique combination of high/low values for Land use indices and Neighborhood Deprivation. Green cells indicate low values and red cells indicate high values. **Given adequate sample size, models that explore interactions between relevant land uses and deprivation can provide further insights.**

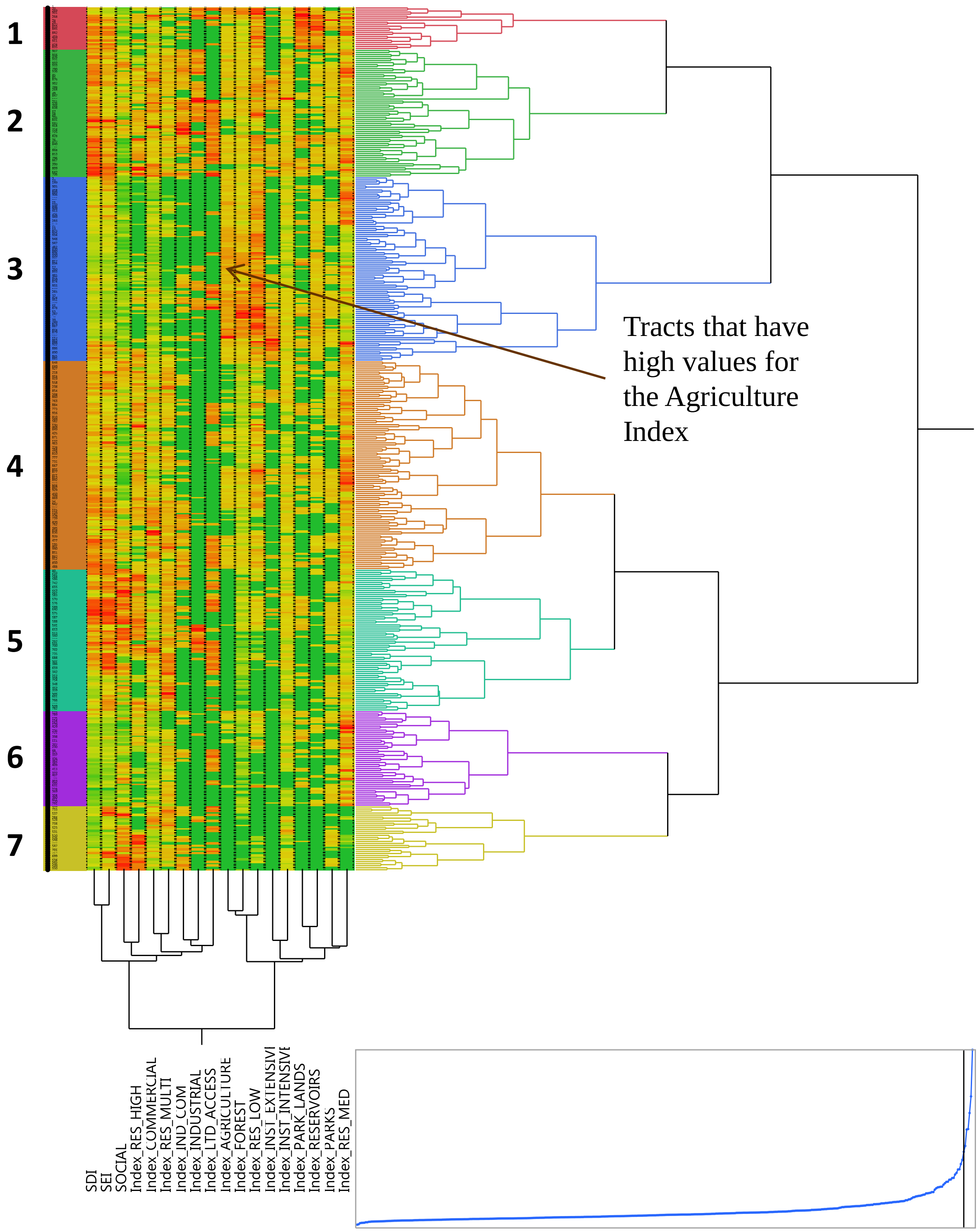


Figure 45. Hierarchical clustering diagram showing unique landscape signatures for tracts with high rates of Lung Cancer mortality

2. ***How is land use mix and spatial distribution (composition and configuration) of landscape components (land use and socioeconomics) associated with health outcomes?***

Two measures of land use diversity were calculated by Patch Analyst— the Shannon Diversity Index (SDI) and the Shannon Evenness Index (SEI). While SDI calculates relative proportions of different land use types in the landscape, SEI also takes into account if the areas are evenly divided among land use types. It provides one value for the entire landscape (i.e. the census tract). Both indices reflect the relative distribution in area between patch types but do not account for spatial distribution. SDI and SEI are used frequently in transportation and physical activity research to evaluate walkability. **Diversity indices were statistically insignificant in all models**.

It is important to note that diversity indices used in walkability research only evaluate a specific set of land use types known to encourage active transportation (commercial, greenspace, residential). Since a number of different health outcomes were being analyzed in this research, a general diversity index was used. This study provides the foundation for developing more specific diversity indices based on the disease outcome of interest and their associations with land uses revealed here.

The importance of configuration was explored through a two-step method. Level 1 land use indices were first constructed from the independent landscape metrics. Through Principal Components Analysis (PCA), the first component was selected to create a single index for each of the 15 selected land use types. The aim was to resolve the multicollinearity seen among landscape metrics of the same land use type. The level 1 indices were used as a first step towards detecting significant land uses.

Level 2 indices represent disaggregated indices for each land use. Here, 3 separate sub-indices were created for each land use type. The level 2 indices measured the domains of Geometry, Shape and Interspersion. Level 2 indices for land use types that were significant from the level 1 analysis were used further for modeling purposes. This method was used specifically to investigate the independent contributions of composition and configuration.

The sub-index “Interspersion” is an explicit measure of spatial arrangement or configuration. For example, the Interspersion/Juxtaposition Index (IJI) which contributes to the Interspersion index is a measure of patch adjacency. IJI approaches 100 when all patch types are equally adjacent to each other. Combined with the Mean Nearest Neighbor and Mean Proximity Index metrics, it provides insights into configuration attributes such as clumping, fragmentation and juxtaposition.

The sub-indices of Geometry and Shape capture measures of both composition and configuration. For example, measures such as number of patches, median patch size and patch size coefficient of variation provide insights into diversity and distribution of patch types within the same land use type. They also appear more meaningful when assessed together rather as separate metrics. The Shape sub-index measures aspects of shape complexity. Combined with the geometry index, it provides insights into land use fragmentation.

The significance of configuration metrics from the model results were mixed. In the case of Agriculture, the Geometry index was dominant and stayed consistently significant in all iterations, across all model types. The same held true for COPD as well. The Interspersion index for Medium Density Residential was occasionally significant for

Lung Cancer. The Interspersion index for Industrial was significant in occasional iterations for COPD. While they didn’t meet the stringent criteria described earlier to be declared as significant findings, it is worth discussing the implications.

First, Agricultural land uses are concentrated in the outer tracts/counties of the study region and are the dominant land use type in those areas. Hence the presence and dominance in the landscape is best measured by the Geometry index. However, Medium Density Residential is a common type of residential land use found in these tracts (also Low Density residential). This is confirmed by signature 3 in fig…..Hence interspersion (adjacency to Agriculture) might be conditionally significant based on context. The literature notes that residents of agricultural communities can be at risk because they work in the farms as well as live close to them.

Second, there are potential attributes of the models themselves to consider in addition to the theoretical explanations. The Neighborhood Deprivation Index is so significant and dominant that it explains most of the variance in the model. The remaining variance is divided between the land use variables and those factors that are unaccounted for in the model. The level 1 index aggregates these variances for each land use type. The interspersion and other configuration measures load prominently on the first component, indicating high correlation with it. These correlations are on par with the composition metrics. However, the level 2 (disaggregated) indices partition the variances further, making it harder to detect land use effects. **Hence, an important consideration is the need to balance theoretical with the mathematical limitations of the statistical models.**

Two recommendations are made for further exploration of composition and configuration. The Hierarchical Clustering and other analyses presented here yield clues as to certain types of juxtapositions that are worth exploring further. Metrics such as the Contrast Weighted Edge Density (CWED) specifically measure adjacencies/juxtapositions based on user-defined weights matrix. Higher weights can be assigned to more/less desirable juxtapositions, based on the problem definition. There were too many potential permutations and combinations of weights to have been explored for the immediate dissertation. Preliminary baselines and particular combinations that vary spatially and with disease type are established through this research and can be used to define relevant contrast weight matrices. Calculating and testing CWED metrics derived from these matrices would be a useful avenue of research in exploring the configuration question further.

The land use indices can be disaggregated further to improve interpretation and policy implications. For example, all the variables that contribute to significant level 2 indices can be used individually in the model to assess independent contributions outside of the index. However, both these techniques would ideally require a much larger sample size. In the case of the multilevel logistic models (explained in the next section), sample size adequacy is important both at the tract and county levels to ensure reliable inference.

3. At what scale do landscape patterns significantly impact human health outcomes?

The tract and county are two nested spatial scales of interest to this research question. The tract represents a local geography (it was also the smallest unit at which health data was available). The county represents the regional geography of interest for modeling purposes. Together, they comprise the 21-county jurisdiction of the Atlanta Regional Commission. Multilevel logistic models were constructed to examine if tract level explanatory variables maintain their significance once county level effects were introduced. **For all disease types, tract level variables found to be significant stayed significant in the presence of county effects. This is potential evidence that land use and socioeconomic factors are locally relevant.**

Regional impacts were captured in the model through county-level random effects. The goal was to test if all variation in health outcomes at the tract level was wholly or partially explained by factors at the county level. Factors such as healthcare access are better assessed at the county scale as they represent more realistic travel thresholds. There are also other measures of county health determinants available through secondary sources such as the County Health Rankings website (RWJF, 2010).

The intercept for each county was allowed to vary (hence random effects). The contextual contribution of each county could be independently assessed this way. Spatial autocorrelation between tracts arising from membership within the same county was also accounted for through this technique. Interesting implications for county level factors were revealed through this analysis. While tract level variables (Land Use and Neighborhood Deprivation Indices) stayed significant, the county intercepts had different magnitudes and directions. Negative intercepts indicate that there are certain protective factors at the county level that reduce the risk for disease. Positive intercepts indicate that there are county level contextual factors that further increase the risk for disease.

Hierarchical Clustering (Figure 46) was again used to investigate potential explanatory variables and their relationships with the positive/negative intercepts. 2010 county-level data from the County Health Rankings database were included in the analysis along with lung cancer mortality rates as a demonstrative example. These included measure of healthcare access (rate of preventable hospital stays, Primary Care Physician rate) as well as measures such as income inequality (GINI). The patterns revealed through the clustering process (Figure43) coincide with the direction of county intercepts. In the case of “healthy” counties (negative intercepts) we see high values for the good factors (red cells) and green values for the bad factors (green cells). The county names labeled in blue represent “healthy” counties, whereas the other group represents “unhealthy” signatures.

While other factors such as %Obese and %Healthy Food are better measured at the tract level, the data is not readily available. From a policy perspective, there is evidence that regional factors such as better healthcare access influence mortality rates. Multilevel models have high sample size requirements, especially if fixed effects are to be reliably estimated at both spatial scales. When sufficient samples sizes are available, county-level fixed effects (landscape metrics, healthcare, and socioeconomics) can be added to further test their statistical significance. The impact of land use planning can be assessed more specifically through this method as well.

In summary, the results of this research indicate that tract-level factors show significant associations with health outcomes, even after the inclusion of county-level effects. However, county-level (regional) effects such as access to healthcare appear to also impact the intensity of health outcomes in addition to tract-level (local) factors. Additional research with county-level fixed effects that include landscape patterns is recommended to disentangle the variance explained by the single county-level effect into its various components. This can further inform evidence-based policy decisions.

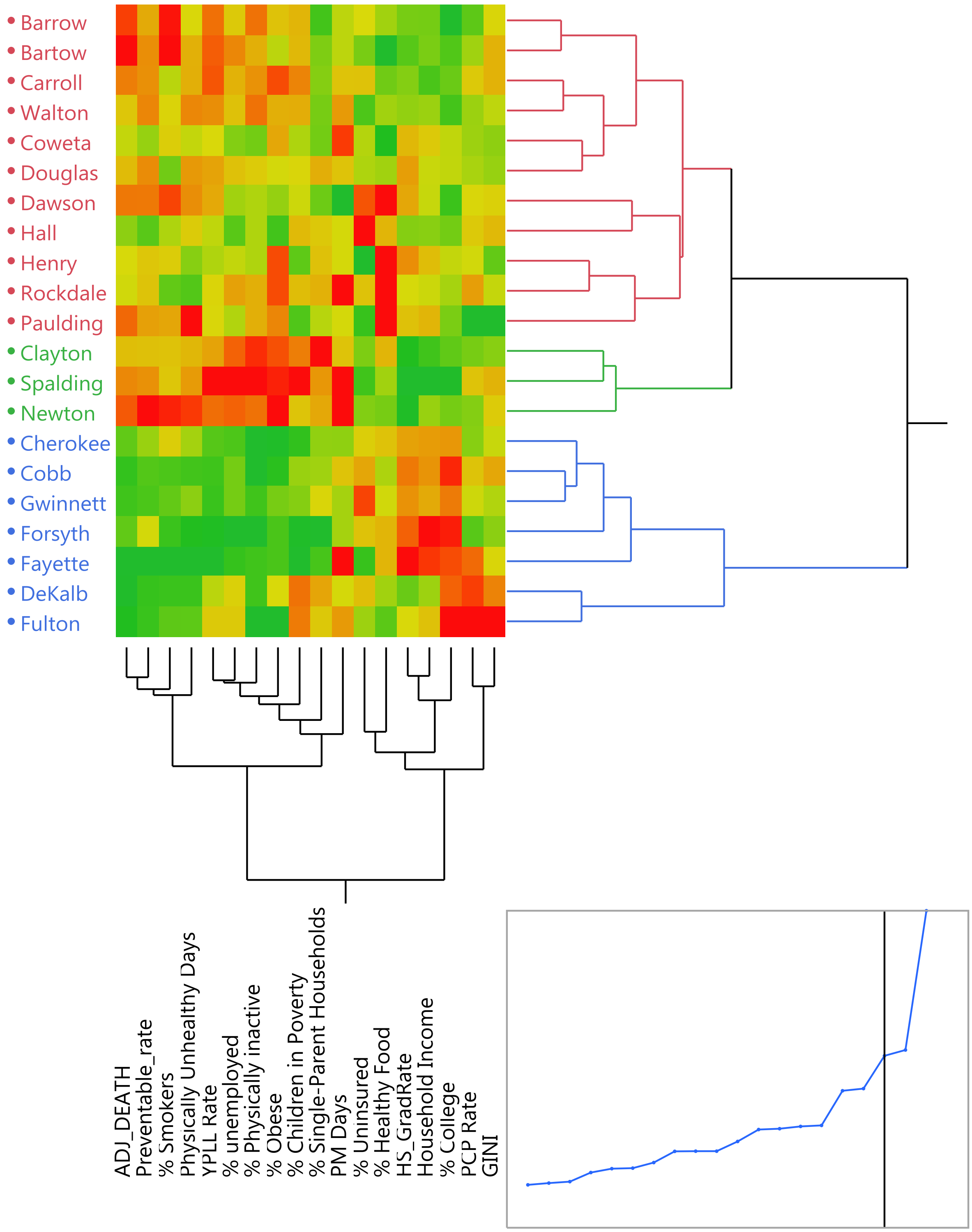


Figure 46. Hierarchical Clustering of county-level lung cancer rates and health determinants

Recommendations for enhancements in future research

Generalizability of the models

As such, the results from this research are applicable only to the Atlanta 21-county study area. These limitations arise from the particular methods for land use classification in the Land Pro dataset as well, use of the Atlanta Metro area population distribution for age standardization as well as the inclusion of fixed effects in the multilevel models.

However, the findings can be made more robust through inclusion of comparable data from different urban areas. However, this is subject to standardization of land use classifications (suggestions provided earlier in the document) and health data across sites. Results from an expanded approach could also provide insights into regional variations in impacts of explanatory variables across metro, micro and other categories of urbanization classifications.

While the results may be specific to the study area, the framework is set up to be easily adaptable to other sites. The dissertation demonstrates a step-by-step process of obtaining raw data, processing/analyzing it and interpreting results.

Alternatives to index construction

The biggest challenge of this research undertaking was the handling of a wide dataset. Furthermore, high correlations existed among the socioeconomic variables and the land use metrics themselves. Most traditional statistical methods are ill-equipped to handle this type of data. Hence, data reduction formed a central component to the methodology. Principal Components Analysis (PCA) was used extensively to construct indices. The method has distinct advantages in a research problem which is exploratory and a priori knowledge of weighting schemes and relationship structures between variables are unknown. PCA provides a data-driven approach to reducing highly correlated data into uncorrelated components. The eigenvectors from these components can be used to aggregate variables into weighted indices. The data-driven approach also highlights its applicability other study areas, as and when data becomes available.

While extremely versatile, PCA still relies on creating linear combinations of data. Hierarchical Clustering is explored as an alternative method to capture non-linear patterns among data. However, this method was not explored to create the indices themselves as they do not provide numerical outputs that can be further analyzed. The PCA and use of the first component worked optimally for the Neighborhood Deprivation Index. This same approach was kept consistent through all the indices constructed. The aim of this study was to provide a baseline for this type of research and to that extent, this approach was efficient and interpretable. It also made a very large dataset more tractable for modeling purposes.

Figure 27 shows evidence of high correlations among landscape metrics of the same land use type. However, the clustering of correlations changes between land use types. This is an indication that landscape metrics behave differently between land use types. All landscape metrics were included in the Level 1 index for each land use type. The Level 2 indices were disaggregated indices built on theoretically-based groupings found in the literature. The same groupings were utilized for all land use types for the sake of consistency of interpretation. However, groupings based on the intrinsic data structures might produce indices that more accurately reflect land use properties. On the flip side, this would create Level 2 indices that were based on abstract constructs and may make interpretation more challenging.

Inclusion of additional explanatory variables and interactions

The dissertation utilized multiple techniques to construct a robust methodology based on the best available data. Various measures were examined in evaluating the quality of the models. Overall, all the models were highly significant and inclusion of the Neighborhood Deprivation Index plays a major role in influencing that significance. The models have a classification accuracy of around 70% and this stays consistent through all model types and iterations. Classification accuracy improves from approximately 68% to 72% when county-level effects are accounted for.

While the aim of this dissertation is focused more on finding associations rather than being predictive, it is important to discuss omitted variables that could potentially improve the classification accuracy of the models. For example, tobacco –smoking is a key predictor in the causal pathway to Lung Cancer. However, the literature reports heterogeneity in tobacco-smoking rates among population groups of interest in the study area. Rates tend to be higher in lower socioeconomic groups whereas rates are lower in agricultural workers. Interactions might also exist when the two types of population characteristics coincide. Additionally, working in agriculture is observed to be a stronger exposure pathway compared to living in closer proximity. A different relationship is observed for Industrial land uses where proximity matters due to stronger environmental externalities. Including a variable that measures occupation status might be useful in explaining misclassified tracts in the central part of the study area. Further examining interactions between socioeconomic factors and land use variables should also be explored. The Hierarchical Clustering diagrams provide insights into these potential interactions. There also appear to be heterogeneity in the interactions themselves based on the signatures observed.

Implications for Urban Planning

The previous sections discuss details of the methodology, potential interpretations of findings for the study area and recommendations for methodological improvement applicable to future research. This final section discusses the larger implications and contributions of the work to the practice of urban planning and health. While the relevance of research lies in its ability to inform practice and policy, it is important to remember that the results presented here do not assume causality in any shape or form. In the hierarchy of epidemiological research, ecological studies similar to this study serve as the foundation for defining and testing new conceptual frameworks. Further refinements to research design are implemented once significant correlations are observed. This research is only envisioned as a starting point in a continuum of investigation regarding landscape patterns and human health.

Current frameworks for incorporating health into urban planning practice

The connection between the urban planning practice and public health fields is emphasized in a few select approaches. Two popular frameworks are discussed are discussed here— Transportation Planning and Health Impact Assessment. While certainly not the only frameworks, the majority of the activities that interface between the two fields occur in these two arenas.

As the MPO for the Atlanta region, the Atlanta Regional Commission provides a really good example through its Regional Transportation Plan (RTP). The RTP is an integral part of the region's Comprehensive Plan. A central tenet of this plan is to create healthy, livable communities and the RTP is considered a key component in achieving this goal. Several elements of the plan address this connection. While they are too numerous to explain in great depth, the planning and health connection is primarily addressed through safety, accessibility, physical activity and air quality. Land use planning is also core to all of these aspects but is primarily a means to the objectives stated above.

Health Impact Assessment (HIA) is another framework that aims to create synergies between urban planning and public health. Firmly embedded in the socioecological model of public health, it is simultaneously holistic and reductionist. The World Health Organization defines HIA as a "practical approach used to judge the potential health effects of a policy, program or project on a population, particularly on vulnerable or disadvantaged groups".

Planning agencies (such as the ARC) and planning organizations (such as the American Planning Association) have used or recommend using HIAs to study the impact of planning decisions on community health. HIAs such as those conducted on the Atlanta Beltline evaluate the multifactorial impact of urban renewal projects on physical activity, social capital and economic development. However, the very definition of HIA treats policies, programs and project as separate entities. In reality, they form a hierarchical continuum and incorporate valuable synergies that may be missed when evaluated separately (Rao and Ross, 2014). They are also often piecemeal, retrospective and not legally mandated. Their impact on decision-making and continued monitoring/evaluation is at best inconsistent.

While both are valuable frameworks as a starting point for including health into urban planning decisions, they have theoretical and methodological limitations. While the transportation frameworks have the ability to inform health, the conversation revolves around measures that enable active transportation, access and air quality as an important externality. Health Impact Assessment is usually applied to individual projects whereas regional applications are starting to be explored (such as the HIA of the PLAN 2040). In summary, we see fragmented methods at different spatial and temporal scales. The framework developed in this dissertation offers a more comprehensive model for the practice of health in urban planning and it does so in the following ways:

Disease Surveillance

In epidemiology, the practice of disease surveillance is to monitor the spread of disease and identify patterns of progression. The main role of disease surveillance is to predict, observe, and mitigate the negative impacts of disease outbreaks. Another key role is to also identify risk factors that contribute to the prevalence, incidence and spread of disease. While health risk behaviors and some environmental factors are tracked for chronic disease surveillance, the built environment is not explicitly included. Moreover, surveillance systems reside in public health agencies.

From the purview of community health, urban planning agencies have the opportunity to offer a much wider conception of health and well-being. The “exposome” concept highlights environmental exposures as an important component of chronic disease causation (Wild, 2005; Buck Louis and Sundaram, 2012) from a life-course perspective.

The research results provide insights into the role of land use and its contribution to the exposome. In particular, it highlights the role of neighborhood deprivation and certain land uses such as agriculture as exposures in disease causation. This puts planning agencies in a primary role in their ability to track and potentially mitigate certain disease conditions.

The notion of land use as exposure reconceives and expands the framework currently used to examine built environment and health relationships. Sprawl, physical activity and other existing transportation frameworks are potentially applicable to a limited spectrum of health outcomes and determinants such as those strongly linked with obesity. However, neighborhood deprivation emerges as a strong predictor of mortality from a variety of causes and can potentially be viewed as a universal indicator.

Urban areas are often classified on gradients of urban form measures. For example, the ARC identifies 6 typologies for characterizing the Atlanta region—Regional Employment Areas, Maturing Neighborhoods, Established Suburbs, Developing Suburbs, Developing Rural and Rural. Again these are density-based frameworks of classifying development patterns and might not completely characterize communities from a health planning perspective. This dissertation introduces other multidimensional classifications such as "Occupation-based communities" or "Deprived communities" which might better capture a wider set of health outcomes and determinants. These are also tied to local land uses in the region and include physical as well as socioeconomic characteristics. An important contribution of this research is the identification of sub-types within these larger typologies. The hierarchical clustering approach creates landscape signatures that indicate spatial heterogeneity and interactions within the broader classification of unhealthy communities.

Specific contributions of this research to healthy communities planning practice are demonstrated through the following scenarios:

Scenario 1- Identification and detection of "leading indicators"

Simply put, Leading Indicators are indicators that signal future events. The various analytical methods utilized in this dissertation present a demonstration of selecting critical variables that have the strongest empirical association with the health outcomes of interest. This research establishes numerical thresholds (baseline) upon which further analysis can be built. Analyzing trends (changes) over time in both socioeconomic and land use variables identified can help develop response models that show time lags between changes in independent variables and resulting changes in dependent variables as well as relative sensitives of the determinants in influencing outcomes.

Hierarchical Clustering, Random Forest and Logistic Regression are cutting-edge classification and variable selection techniques that represent well-established practices in industry. Applying these techniques to temporal data can result in the selection of leading indicators incorporated into a signaling system that prospectively portends changes in health outcomes based on changes in urbanization patterns.

In the era of big data, such data intensive techniques are entirely feasible. The research framework presented here utilizes several readily available data sources and free softwares. Additionally, planning agencies are beginning to develop internal capacity for sophisticated data analytics and modeling.

Scenario 2- Creating risk maps

Incorporating model improvements suggested in previous sections can result in the creation of predictive risk maps that can further inform practice and policy-making. Overcoming the limitations of cross-sectional research as presented in this study would also greatly enhance the potential for predictive modeling. Understanding time lags between land use changes and their corresponding impact on changes in health outcomes is an important longitudinal aspect that must be considered. Logistic Regression is a powerful regression technique that calculates probabilities for disease based on a set of environmental predictors. These probabilities, when mapped, help visualize risk surfaces spatially.

One of the most significant takeaways from this dissertation is the understanding that different diseases have different epidemiological mechanisms and thus need different sets of metrics that link them to environmental exposures. From a surveillance perspective, the research executed here lays the framework for a comprehensive, cohesive and holistic approach for synthesizing data, theory and methods in a way that allows planning organizations to monitor disease as a consequence of built and social environments.

This contribution is echoed through the literature when similar approaches are used for remote sensing and infectious disease (Dambach et al, 2012; Midekisa et al, 2012) as well as socioeconomic patterns and health outcomes. It also demonstrates that a "one size fits all" approach might not be the most appropriate one. It adds a new way of conceptualizing healthy/unhealthy communities and thus significantly expands the built environment-health research. Indeed, a systems-based surveillance structure would be able to mesh and utilize all of these frameworks as appropriate.

Regional and Local Planning

Another significant contribution of this dissertation to regional planning is the creation of a multilevel, multidimensional framework that has applicability for several regional planning purposes. Conceptually, the framework served as a bridge between Urban Planning, Social Epidemiology, Spatial Epidemiology and Landscape Ecology.

Landscape metrics provide a more holistic characterization of the land use environment, encompassing composition and configuration. The utility of these metrics have been theoretically discussed in several disciplines. Their merits have often been proposed as antidotes to methodological shortcomings in transportation, health and environmental planning literature alike. Their association with ecosystem quality has been well established in the literature. Landscape metrics have also been used to characterize sprawl through changes in land cover pattern. Fragmentation, particularly of natural areas is the most significant finding of this stream of research. Both types of research present regional applications of using landscape metrics. However, they rarely proceed beyond descriptive techniques to validate these findings against anthropocentric outcomes of interest.

The study presented here brings this approach full circle by validating it against human health outcomes (mortality rates). It also demonstrates the applicability of using landscape metrics to evaluate local environments (census tracts). It creates a central platform or common language, simultaneously applicable to Transportation Planning and Sustainable Land Use Planning. The Neighborhood Deprivation Index further permits the incorporation of equity interests across all thematic areas.

As described earlier, the most daunting task in this dissertation was the whittling down of a vast number of metrics to a key set of indicators. Several methods were utilized to achieve this goal. Principal Components Analysis and Logistic Regression, along with data mining techniques such as Random Forest were used for variable selection and classification. Multilevel Regression was used to test if county-level effects nullified the significance of local tract-level variables.

With the increasing uptake of big data in planning and public policy, this project provides an end-to-end demonstration of how such datasets can be processed, analyzed and interpreted. Planning agencies such as the ARC continue to build their in-house analytics and research capabilities. This research is timely in that context. The research also demonstrates a synthesis of diverse data types. From personal experience with the American Planning Association in advising their Plan4Health grantees and other project experience at The American Cancer Society, data acquisition, analysis and synthesis present the biggest challenges for local community planning efforts.

Scalability is a very important property of landscape metrics. A brief discussion on the shortcomings of other frameworks such as sprawl with respect to this property is valuable to this discussion. For example, the sprawl index constructed by Ewing and Hamidi (2014) consists of four dimensions— Density, Mixed Use, Centering and Street Accessibility. While Density, Mixed Use and Street Accessibility might be rationalized at the tract level, the Centering dimension holds little validity at the local level. This is consonant with travel elasticities and travel behavior. The authors specifically point out this shortcoming in the methodology. It is also clear that these measures are all derived from the transportation literature.

Undoubtedly, environments that encourage active transportation are important in public health prevention for conditions such as obesity. However, they are limited in their utility for investigating other health outcomes. Conversely, landscape metrics hold their validity, both at local and regional scales. At the local scale, composition and configuration of individual patches as well as their adjacencies can be used to characterize environmental quality and perceptual characteristics. Additionally, the presence/absence of desirable/undesirable land uses can also be evaluated. Shannon diversity indices also provide collective information such as those used in the land use mix component of the sprawl index. At the regional scale, they help evaluate landscape patterns more relevant to ecosystem processes.

The utility of a methodology that is able to create scalable metrics that can evaluate land uses separately and collectively is demonstrated through the modeling results. In all likelihood, sprawl metrics would not have picked up the association between agriculture and lung cancer risk (this is also demonstrated by the paper by Berrigan et al, 2014). The sprawl indices do not work for areas that are less urban. In the presence of additional data, it can be hypothesized within reason that landscape pattern metrics can be used in rural studies as well as in understanding regional ecosystem services and its connection with human health. This is a promising arena for future research and practice applications.

While the results are applicable only to the study area, the models provide the most important takeaways for local land use planning. Unlike other urban planning and health studies, this dissertation began with an inquiry into the most important cause of mortality in the region. This was considered an important first step in creating an approach that is relevant to the region under investigation. The level 1 and level 2 land use indices permitted a simultaneously separate and combined approach to investigating the relationships between land use types and health. They also helped eliminate several insignificant metrics and land use types, creating a more succinct dataset for confirmatory modeling and inference.

The Agriculture Index was consistently significant for Lung Cancer and COPD mortality and the Neighborhood Deprivation Index was highly significant across all four disease outcomes (Lung Cancer, COPD, Heart Disease and Diabetes). Other land use indices for Medium Density residential and Industrial were occasionally significant but provide insights into complementary adjacencies between land uses from a health standpoint. There is evidence that both composition and configuration are conditionally significant. However, this aspect needs to be explored further by using individual metrics rather than land use indices.

Specific land use effects could not be investigated from a regional (county) perspective due to limitations of sample size. However, the multilevel models revealed two important conclusions for land use planning. Local land use indices maintained their significance despite adding county-level effects. This suggests that local land use planning does matter since all variation in the dependent variable was not accounted for by the county-level effects. However, the magnitude and direction of the intercepts provide indications that certain regional properties of counties can impact local disease outcomes. The Hierarchical Clustering diagram sorts the healthy and unhealthy counties based on a series of county level characteristics. While the signature for healthcare access (# preventable hospital admissions, PCP rate) was clean, the signature for the GINI index (income inequality indicating socioeconomic patterns) was unclear. Counties with positive intercepts (increase disease risk) were associated with a lower PCP rate and vice-versa.

From an intervention standpoint, both local and regional land use planning are relevant. For example, the local environment becomes highly significant in dense but deprived communities in the central part of the Atlanta region. In addition to land use, finer grain considerations that reflect food environments and other subtleties are highlighted from existing literature. The distribution of metabolic diseases such as Heart Disease and Diabetes reflect this pattern. On the other hand, agricultural communities show associations with specific land use types. These are located along the fringes of urban areas, often next to forest and other natural areas that face maximum pressure from development patterns. Thus, outer metro areas might require additional land use considerations both from a view point of human and ecosystem health.

The analysis also provides insights into the appropriate locations for local and regional healthcare needs. For example, agricultural communities can benefit from improved lung cancer screening interventions whereas deprived communities might benefit from tobacco-cessation programs. Physical activity and nutrition education might also become an important component of prevention strategies for deprived communities with denser populations. The variations in landscape signatures can also be further analyzed to develop strategic, context-relevant interventions.

Conclusions

Overall, this dissertation was successful in demonstrating a comprehensive framework for investigating the relationships between chronic disease outcomes and land use patterns. Landscape metrics are already being utilized for ecosystems and infectious disease se research (Landscape Epidemiology). It is theoretically proposed as a superior method for transportation research as well (Hess et al, 2001). The work accomplished here expanded its scope and utility for chronic disease research. It is an exciting possibility to envision that a single measurement framework can simultaneously be utilized to accomplish every dimension of a Regional Planning organization's transportation and sustainable land use planning needs.

More specifically, model improvements and layering of data can make it useful in analyzing and monitoring changes in a wide spectrum of disease types in response to changes in urban patterns involving both physical and socioeconomic characteristics. Ideally speaking, a framework that combines the explanatory power of landscape metrics along with measures such as sprawl captures the complexity of the built environment and the various mechanisms through which it influences health.

The work and results are also timely for my future research goals as a researcher at The American Cancer Society (ACS). Public health organizations are increasingly realizing the importance of healthy communities' frameworks in their research and practice. I hope to use this work to contribute towards their mission of cancer prevention.

REFERENCES

[1]. Acevedo-Garcia, D., Lochner, K., Osypuk, T., & Subramanian, S. (2003). Future directions in residential segregation and health research: a multilevel approach. American Journal of Public Health, 93(2), 215-221.

[2]. Alavanja MC, Dosemeci M, Samanic C, Lubin J, Lynch CF, Knott C, Barker J, Hoppin JA, Sandler DP, Coble J, Thomas K, Blair A. (2004). Pesticides and lung cancer risk in the agricultural health study cohort. Am J Epidemiol., 60(9):876-885.

[3]. Alavanja, M. R., Ward, M. H., & Reynolds, P. (2007). Carcinogenicity of agricultural pesticides in adults and children. Journal of Agromedicine, 12(1), 39-56.

[4]. Alavanja, M. R., Ward, M. H., & Reynolds, P. (2007). Carcinogenicity of agricultural pesticides in adults and children. Journal Of Agromedicine, 12(1), 39-56.

[5]. Alberti, M. (2005). The Effects of Urban Patterns on Ecosystem Function. International Regional Science Review,28(2), 168-192. doi:10.1 177/0160017605275160

[6]. Alberti, M. (2007). Ecological Signatures: The Science of Sustainable Urban Forms. Places: Forum of Design for the Public Realm, 19(3), 56-60.

[7]. Alberti, M. M. (2008). Advances in urban ecology [electronic resource]: integrating humans and ecological processes in urban ecosystems / Marina Alberti. New York : Springer, c2008

[8]. Alberti, M., & Marzluff, J. M. (2004). Ecological resilience in urban ecosystems: Linking urban patterns to human and ecological functions. Urban Ecosystems, 7(3), 241-265.

[9]. Alberti, M., Marzluff, J. M., Bradley, G., Ryan, C., Shulenberger, E., & Zumbrunnen, C. (2003). Integrating Humans into Ecology: Opportunities and Challenges for Studying Urban Ecosystems. Bioscience, 53(12), 1169-1179.

[10]. American Cancer Society Cancer Action Network (ACS CAN).2009. Cancer Disparities: A Chartbook. http://action.acscan.org/site/DocServer/cancer-disparities-chartbook.pdf.

[11]. Anselin, A.,Lozano, N., and Kochinsky, J. (2006). Rate Transformations and Smoothing. Spatial Analysis Laboratory Department of Geography University of Illinois, Urbana-Champaign. https://geodacenter.asu.edu/pdf/smoothing_06.pdf.

[12]. Anselin, L., I. Syabri and Y Kho. (2005). GeoDa : An Introduction to Spatial Data Analysis. Geographical Analysis 38(1), 5-22.

[13]. Arbex, M. A., de Souza Conceição, G. M., Cendon, S. P., Arbex, F. F., Lopes, A. C., Moysés, E. P., … Braga, A. L. F.. (2009). Evidence-based public health policy and practice: Urban air pollution and chronic obstructive pulmonary disease-related emergency department visits. Journal of Epidemiology and Community Health, 63(10), 777–783. Retrieved from http://www.jstor.org/stable/20721057.

[14]. Barnes, P.J., Adcock, I.M. (2011). Chronic Obstructive Pulmonary Disease and Lung Cancer: A Lethal Association. American Journal of Respiratory and Critical Care Medicine, 184(8), pp. 866-867. doi: 10.1164/rccm.201108-1436ED.

[15]. Barrington, W. E., Stafford, M., Hamer, M., Beresford, S. A., Koepsell, T., & Steptoe, A. (2014). Neighborhood socioeconomic deprivation, perceived neighborhood factors, and cortisol responses to induced stress among healthy adults. Health & Place,27120-126. doi:10.1016/j.healthplace.2014.02.001

[16]. Berkman LF, Macintyre S.(1997). The measurement of social class in health studies: old measures and new formulations. In: Kogevinas M, Pearce N, Susser M, Boffetta P, eds. Soc Inequalities Cancer. Lyon, France: International Agency for Research on Cancer; pg.51–64.

[17]. Berrigan, D. and McKinno,R.A. (2008). Built Environment and Health. Prev Med, 47(3): 239–240.

[18]. Berrigan, D., Tatalovich, Z., Pickle, L. W., Ewing, R., & Ballard-Barbash, R. (2014). Urban sprawl, obesity, and cancer mortality in the united states: Cross-sectional analysis and methodological challenges. International Journal of Health Geographics, 13(1), 3. doi:http://dx.doi.org.prx.library.gatech.edu/10.1186/1476-072X-13-3.

[19].Berrigan, David, Zaria Tatalovich, Linda W. Pickle, Reid Ewing, and Rachel Ballard-Barbash. 2014. "Urban sprawl, obesity, and cancer mortality in the United States: cross-sectional analysis and methodological challenges." International Journal Of Health Geographics 13, no. 1: 2-27. Academic Search Complete, EBSCOhost (accessed May 31, 2015).

[20].Biau,G., Scornet, E., (2015). A Random Forest Guided Tour. Cornell University

[21].Blair A, Freeman LB (2009) Epidemiologic studies in agricultural populations: observations and future directions. J Agromedicine 14(2):125–131

[22].Blanc, P. D. (2012). Occupation and COPD: A brief review. The Journal Of Asthma, 49(1), 2-4. doi:10.3109/02770903.2011.611957

[23].Bodea, T. D., Garrow, L. A., Meyer, M. D., & Ross, C. L. (2009). Socio-demographic and built environment influences on the odds of being overweight or obese: the Atlanta experience. Transportation Research Part A: Policy and Practice, 43(4), 430-444. doi:10.1016/j.tra.2008.11.009.

[24].Boulos, D. N., Kamel, Ghali, R. R., Ibrahim, E. M., Boulos, M. N., Kamel, & Abdelmalik, P. (2011). An eight-year snapshot of geospatial cancer research (2002-2009): Clinico-epidemiological and methodological findings and trends. Medical Oncology,28(4), 1145-1162. doi:http://dx.doi.org.prx.library.gatech.edu/10.1007/s12032-010-9607-z

[25].Brackbill, R.M., L.L.Cameron, and V.Behrens. 1994. Prevelance of chronic diseases and impairments among U.S. farmers, 1986-1990. Am. J. Epidemiol. 139(11):1055-65.

[26].Bradman A, Eskenazi B, Barr DB, Bravo R, Castorina R, Chevrier J, Kogut K, Harnly ME, McKone TE. (2005). Organophosphate urinary metabolite levels during pregnancy and after delivery in women living in an agricultural community. Environ Health Perspect.,113(12):1802-1807.

[27].Brown, Pijanowski, & Duh. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. Journal of Environmental Management, 59(4), 247-263.

[28]. Buck Louis, G. M., & Sundaram, R. (2012). Exposome: Time for Transformative Research. Statistics in Medicine, 31(22), 10.1002/sim.5496. http://doi.org/10.1002/sim.5496

[29]. Businelle, M. S., Kendzor, D. E., Reitzel, L. R., Costello, T. J.. Cofta-Woerpel, L., Li, Y,. Mazas, C. A., Vidrine, J. I., Cinciripini, P. M., Greisinger, A, J., Wetter, D. W. (2010). Mechanisms linking socioeconomic status to smoking cessation: A structural equation modeling approach. Health Psychology, 29(3), 262-273. http://dx.doi.org/10.1037/a0019285

[30]. C. Raherison, & P-O Girodet. (2009). Epidemiology of COPD. European Respiratory Review, 18(114), 213-221.

[31]. CDC. (2007). Respiratory Disease in Agricultural Workers: Mortality and Morbidity Statistics. Department of Health and Human Services. http://www.cdc.gov/niosh/docs/2007-106/pdfs/2007-106.pdf, Accessed 04/01/2016.

[32]. Cervero, R., & Murakami, J. (2010). Effects of built environments on vehicle miles traveled: Evidence from 370 U.S. metropolitan areas. Environment and Planning A, 42(2), 400–418.

[33]. Charnita M. Zeigler-Johnson, Ann Tierney, Timothy R. Rebbeck, and Andrew Rundle, "Prostate Cancer Severity Associations with Neighborhood Deprivation," Prostate Cancer, vol. 2011, Article ID 846263, 9 pages, 2011. doi:10.1155/2011/846263.

[34]. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. New Engl. J. Med. 2008;358:2249–2258.

[35]. Christian, H. E., Bull, F. C., Middleton, N. J., Knuiman, M. W., Divitini, M. L., Hooper, P. Giles-Corti, B. (2011). How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study. International Journal of Behavioral Nutrition & Physical Activity, 8(1), 55-66. doi: 10.1186/1479-5868-8-55.

[36]. Comber A. J., R. A. Wadsworth & P. F. Fisher. (2008). Using semantics to clarify the conceptual confusion between land cover and land use: the example of 'forest', Journal of Land Use Science, 3:2-3, 185-198

[37]. Corburn, J. (2005). Urban planning and health disparities: Implications for research and practice. Planning Practice & Research, 20(2), 111-126. doi:10.1080/02697450500414652

[38]. Corburn, J. (2007a). Urban land use, air toxics and public health: Assessing hazardous exposures at the neighborhood scale. Environmental Impact Assessment Review, 27(2), 145-160. doi: 10.1016/j.eiar.2006.10.004.

[39]. Corburn, J. (2007b). Reconnecting with Our Roots: American Urban Planning and Public Health in the Twenty-First Century. Urban Affairs Review, 42(5), 688-713.

[40]. Corburn, J. (2012). Reconnecting Urban Planning and Public Health. The Oxford Handbook Of Urban Planning, doi:10.1093/oxfordhb/9780195374995.013.0020

[41]. Cornelius, C., Schoeters, G., Kellen, E., Buntinx, F., & Zeegers, M. (2009). Development of a GIS-based indicator for environmental pesticide exposure and its application to a Belgian case-control study on bladder cancer. International Journal of Hygiene and Environmental Health, 212(2), 172-85.

[42]. Corvalan, C., Hales, S., McMichael, A. et al. (2005). Ecosystems and human well-being: health synthesis: a report of the Millennium Ecosystem Assessment. World Health Organization

[43]. Coutts, C. J. (2010). Public Health Ecology. Journal Ofof Environmental Health, 72(6), 53-55.

[44]. Coutts, C., & Taylor, C. (2011). Putting the capital "E" environment into ecological models of health. Journal Ofof Environmental Health, 74(4), 26-29.

[45]. Crews, K. A., & Peralvo, M. F. (2008). Segregation and Fragmentation: Extending Landscape Ecology and Pattern Metrics Analysis to Spatial Demography. Population Research And Policy Review, (1), 65. doi:10.2307/41217937.

[46]. Cromley. E.K and McLafferty, S.L. (2012). “Analyzing Spatial Clustering of Health Events” in GIS and Public Health, Guilford Publications, pg.150-182.

[47]. Cushman, S. A., McGarigal, K., & Neel, M. C. (2008). Parsimony in landscape metrics: Strength, universality, and consistency. Ecological Indicators, 8691-703. doi:10.1016/j.ecolind.2007.12.002

[48]. Cushman, S. A., Neel, M. C., & McGarigal, K. (2008). Parsimony in landscape metrics: Strength, universality, and consistency [electronic resource]. Ecological Indicators, 8(5), 691-703.

[49]. Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. J. Health Econ. 2010;29(1):1–28.

[50]. Dambach, P., Machault, V., Lacaux, J., Vignolles, C., Sie, A., & Sauerborn, R. (2012). Utilization of combined remote sensing techniques to detect environmental variables influencing malaria vector densities in rural West Africa. International Journal of Health Geographics, 11, 8.

[51]. Dean, J., van Dooren, K., & Weinstein, P. (2011). Does biodiversity improve mental health in urban settings? Medical Hypotheses, 76(6), 877.

[52]. Diez Roux, A.V. (2001). Investigating Neighborhood and Area Effects on Health. American Journal of Public Health, 91(11), pp. 1783-1789. doi: 10.2105/AJPH.91.11.1783

[53]. Duncan, M., Winkler, E., Sugiyama, T., Cerin, E., duToit, L., Leslie, E., & Owen, N. (2010). Relationships of Land Use Mix with Walking for Transport: Do Land Uses and Geographical Scale Matter? Journal of Urban Health, 87(5), 782-795. doi: 10.1007/s11524-010-9488-7.

[54]. Durham, A., & Adcock, I. (2015). The relationship between COPD and lung cancer. Lung Cancer (Amsterdam, Netherlands),90(2), 121-7.

[55]. Earle-Richardson, G., Scribani, M., Scott, E., May, J., & Jenkins, P. (2015). A comparison of health, health behavior, and access between farm and nonfarm populations in rural New York state. The Journal Of Rural Health, 31(2), 157-164. doi:10.1111/jrh.12098.

[56]. Earle-Richardson, G., Scribani, M., Scott, E., May, J., & Jenkins, P. (2015). A comparison of health, health behavior, and access between farm and nonfarm

populations in rural New York state. The Journal Of Rural Health: Official Journal Of The American Rural Health Association And The National Rural Health Care Association, 31(2), 157-164. doi:10.1111/jrh.12098.

[57]. Eid, J., Overman, H. G., Puga, D., & Turner, M. A. (2008). Fat city: Questioning the relationship between urban sprawl and obesity.Journal Of Urban Economics, 63385-404. doi:10.1016/j.jue.2007.12.002

[58]. Elkie, P.C., Rempel, R. S., and Carr, A. P. (1999). Patch Analyst User's Manual: A Tool for Quantifying Landscape Structure. Ont. Min.Natur. Resour. Northwest Sci. & Technol. Thunder Bay, Ont. TM–002.

[59]. Elliott, P., & Wartenberg, D. (2004). Spatial Epidemiology: Current Approaches and Future Challenges. Environmental Health Perspectives, (9). 998.

[60]. Ewing, R. H., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. Journal Of The American Planning Association, 76(3), 265-294.

[61]. Ewing, R., & Handy, S. (2009). Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. Journal of Urban Design, 14(1), 65-84. doi:10.1080/13574800802451155

[62]. Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and Measuring Urban Design Qualities Related to Walkability. Journal of Physical Activity & Health, 3S223.

[63]. Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., and Raudenbush, S. (2003). Relationship between Urban Sprawl and Physical Activity, Obesity and Morbidity. American Journal of Health Promotion, 18(1), p.47-57.

[64]. Ewing,R and Hamidi, S. (2014). Measuring Urban Sprawl and Validating Sprawl Measures. National Institutes of Health and Smart Growth America.

[65]. Factor, R.)., Awerbuch, T.)., & Levins, R.). (2013). Social and land use composition determinants of health: Variability in health indicators. Health And Place, 2290-97. doi:10.1016/j.healthplace.2013.04.001

[66]. Feng, J., Glass, T.,Curriero, F.C., Stewart, W.F. and Schwartz, B.S. (2010). The built environment and obesity: A systematic review of the epidemiological evidence. Health and Place, 16, p.175-190.

[67]. Fern´andez-Delgado, M., Cernadas, E., Barro, S., Amorim, D. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? Journal of Machine Learning Research, 15, p.3133-3181.

[68]. Flatt, William P. "Agriculture in Georgia: Overview." New Georgia Encyclopedia. 09 December 2015.

[69]. Frank, L. D., & Engelke, P. (2005). Multiple Impacts of the Built Environment on Public Health: Walkable Places and the Exposure to Air Pollution. International Regional Science Review, 28(2), 193-216. doi: 10.1177/0160017604273853

[70]. Frank, L. D., & Engelke, P. O. (2001). The built environment and human activity patterns: exploring the impacts of urban form on public health. Journal of Planning Literature, 16(2), 202-218.

[71]. Frank, L. D., Andresen, M. A., & Schmid, T. L. (2004). Obesity relationships with community design, physical activity, and time spent in cars. American Journal of Preventive Medicine, 27(2), 87-96. doi: http://dx.doi.org/10.1016/j.amepre.2004.04.011

[72]. Frank, L. D., Sallis, J. F., Conway, T. L., Chapman, J. E., Saelens, B. E., & Bachman, W. (2006). Many Pathways from Land Use to Health. Journal of the American Planning Association, 72(1), 75-87.

[73]. Frenkel, A., & Ashkenazi, M. (2008). Measuring urban sprawl: How can we deal with it? Environment and Planning B: Planning and Design, 35(1), 56-79.

[74]. Gale, S.)., Magzamen, S.)., Tager, I.)., & Radke, J.). (2011). Crime, neighborhood deprivation, and asthma: A GIS approach to define and assess neighborhoods. Spatial And Spatio-Temporal Epidemiology, 2(2), 59-67. doi:10.1016/j.sste.2011.01.001

[75]. George Galster , Royce Hanson , Michael R. Ratcliffe , Harold Wolman , Stephen Coleman & Jason Freihage (2001) Wrestling Sprawl to the Ground: Defining and measuring an elusive concept, Housing Policy Debate, 12:4, 681-717, DOI:

[76]. Georgia Department of Public Health. (2012). Selected Sociodemographic Characteristics by Demographic Cluster. https://oasis.state.ga.us/gis/demographiccluster/documents/EASIDemographics2011.pdf

[77]. Gottdenker, N., Streicker, D., Faust, C., & Carroll, C. (2014). Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence. Ecohealth.

[78]. Gregson, J. (2011). Poverty, Sprawl, and Restaurant Types Influence Body Mass Index of Residents in California Counties. Public Health Reports, 126141-149.

[79]. Griffin, B. A., Eibner, C., Bird, C. E., Jewell, A., Margolis, K., Shih, R., & ... Escarce, J. J. (2013). The relationship between urban sprawl and coronary heart disease in women. Health & Place, 2051-61. doi:10.1016/j.healthplace.2012.11.003

[80]. Gustafson, A., Lewis, S., Perkins, S., Damewood, M., Buckner, E., Vail, A., & ... Jilcott-Pitts, S. B. (2013). Association Between the Retail Food Environment, Neighborhood Deprivation, and County-Level Dietary Outcomes Among Supplemental Nutrition Assistance Program–Education (SNAP-Ed) Recipients in Kentucky, 2010–2011. Journal Of Hunger & Environmental Nutrition, 8(3), 362-377. doi:10.1080/19320248.2013.816993.

[81]. Hale,N., Probst,J.C. and Robertson, A. (2015). Area Deprivation is Higher Among Rural Counties- but Not All Rural Counties are deprived. South Carolina Rural Health Research Center. http://rhr.sph.sc.edu/report/(14-31)Area_deprivation_higher_among_rural_Counties.pdf

[82]. Handy, S. L., Xinyu, C., & Mokhtarian, P. L. (2008). The Causal Influence of Neighborhood Design on Physical Activity within the Neighborhood: Evidence from Northern California. American Journal of Health Promotion, 22(5), 350-358.

[83]. Hankey, S., Marshall, J. D., & Brauer, M. (2012). Health Impacts of the Built Environment: Within-Urban Variability in Physical Inactivity, Air Pollution, and Ischemic Heart Disease Mortality. Environmental Health Perspectives, 120(2), 247-253.

[84]. Harper S, Lynch J. 2005. Methods for Measuring Cancer Disparities: Using Data Relevant to Healthy People 2010 Cancer-Related Objectives. NCI Cancer Surveillance Monograph Series, Number 6. Bethesda, MD: National Cancer Institute,. NIH Publication No. 05-5777.

[85]. Herold, M., Scepan, J., & Clarke, K. C. (2002). The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. Environment & Planning A, 34(8), 1443-1458.

[86]. Hess, P.M., Moudon, A.V. and Logsdon, M. G. (2001). Measuring Land Use Patterns for Transportation Research. Transportation Research Record: Journal of the Transportation Research Board, 1780(1), p.17-24.

[87]. Hiscock, R., Bauld, L., Amos, A., Fidler, J. A. and Munafò, M. (2012), Socioeconomic status and smoking: a review. Annals of the New York Academy of Sciences, 1248: 107–123. doi: 10.1111/j.1749-6632.2011.06202.x

[88]. Hope, D. D., McIntyre, N. E., & Knowles-Yanez, K. K. (2000). Urban ecology as an interdisciplinary field: differences in the use of "urban" between the social and natural sciences. Urban Ecosystems, 4(1), 5.

[89]. Houghton, A. M. (2013). Mechanistic links between COPD and lung cancer. Nature Reviews. Cancer, 13(4), 233-245. doi:10.1038/nrc3477

[90]. Hu, H., Cho, J., Huang, G., Wen, F., Choi, S., Shih, M., & Lightstone, A. S. (2014). Neighborhood environment and health behavior in Los Angeles area. Transport Policy, 3340-47. doi:10.1016/j.tranpol.2014.02.004

[91]. Jackson, R., Dannenberg, A., & Frumkin, H. (2013). Health and the built environment: 10 years after. American Journal of Public Health, 103(9), 1542-1544. doi:10.2105/AJPH.2013.301482

[92]. Jackson, R., Harp, T., & Wright, T. (2002). Land use planning: why public health must be involved. The Journal of Law, Medicine & Ethics: A Journal of The American Society of Law, Medicine & Ethics, 30(3 Suppl), 70-74.

[93]. Jacquez, GM. 2008. “Spatial Cluster Analysis”. Chapter 22 in The Handbook of Geographic Information Science, S. Fotheringham and J. Wilson (Eds.). Blackwell Publishing, pages 395-416.

[94]. James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer, New York, NY.

[95]. Jensen, J. (2005). Introductory digital image processing : A remote sensing perspective (3rd ed., Prentice Hall series in geographic information science). Upper Saddle River, N.J.: Prentice Hall.

[96]. Ji, W., Ma, J., Twibell, R., & Underhill, K. (2006). Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics. Computers, Environment And Urban Systems, 30(Location Based Services), 861-879. doi:10.1016/j.compenvurbsys.2005.09.002.

[97]. Jianguo, L., Dietz, T., Carpenter, S. R., Folke, C., Alberti, M., Redman, C. L., & ... Provencher, W. (2007). Coupled Human and Natural Systems. AMBIO - A Journal of The Human Environment, 36(8), 639-649.

[98]. John R. Logan, Zengwang Xu, and Brian Stults. 2012. “Interpolating US Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database” Professional Geographer, 66,3: 412-420.

[99]. Johnson, S., & Marko, J. (2008). Designing Healthy Places: Land use planning and public health. Environments: A Journal of Interdisciplinary Studies, 35(3), 9-19.

[100]. Jolliffe, I. 2014. Principal Component Analysis. Wiley StatsRef: Statistics Reference Online.

[101]. Jones RR, Barone-Adesi F, Koutros S, Lerro CC, Blair A, Lubin J, Heltshe SL, Hoppin JA, Alavanja MC, Beane Freeman LE. (2015). Incidence of solid tumors among pesticide applicators exposed to the organophosphate insecticide diazinon in the Agricultural Health Study: an updated analysis. Occup Environ Med., 72(7):496-503. doi: 10.1136/oemed-2014-102728.

[102]. Joshu, C.E., Boehmer, T.K., Brownson, R.C. and Ewing, R. (2008). Personal, neighborhood and urban factors associated with obesity in the United States. Journal of Epidemiology and Community Health, 62, p.202-208.

[103]. Kaczynski, A. T., Johnson, A. J., & Saelens, B. E. (2010). Neighborhood land use diversity and physical activity in adjacent parks. Health & Place, 16(2), 413-415. doi: 10.1016/j.healthplace.2009.11.004

[104]. Kartaloglu, Z. (2013). Socioeconomic Status and Chronic Obstructive Pulmonary Disease. TAF Preventive Medicine Bulletin, 12 (1), 87-96. Turkish. doi:10.5455/pmb.1-1331198324.

[105]. Kearns, R., & Moon, G. (2002). From medical to health geography: novelty, place and theory after a decade of change. Progress in Human Geography, 26(5), 605-625. doi:10.1191/0309132502ph389oa.

[106]. Kim JH, Lee C, Olvara NE, Ellis CD. (2013). The Role of Landscape Spatial Patterns on Obesity in Hispanic Children Residing in Inner-City Neighborhoods. J Phys Act Health. 2013 Dec 31. [Epub ahead of print].

[107]. Kirkhorn, S. R., & Garry, V. F. (2000). Agricultural lung diseases.Environmental Health Perspectives, 108(Suppl 4), 705–712.

[108]. Ko, F. W. S., Hui, D.S.C., (2012). Air pollution and chronic obstructive pulmonary disease. Respirology , 17(3), p.395-401.

[109]. Koshiol, J., Rotunno, M., Consonni, D., Pesatori, A., De Matteis, S., Goldstein, A., . . . Vij, N. (2009). Chronic Obstructive Pulmonary Disease and Altered Risk of Lung Cancer in a Population-Based Case-Control Study (COPD and Lung Cancer).PLoS ONE, 4(10), E7380.

[110]. Kostova, D. (2011). Can the built environment reduce obesity the impact of residential sprawl and neighborhood parks on obesity and physical activity. Eastern Economic Journal, 37(3), 390-402. doi:10.1057/eej.2009.47

[111]. Krieger N, Williams DR, Moss NE. (1997). Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health, 18, p.341–378.

[112]. Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider?. Social Science & Medicine, 39887-903. doi:10.1016/0277-9536(94)90202-X

[113]. Kupfer, J. A. (2012). Landscape ecology and biogeography: Rethinking landscape metrics in a post-FRAGSTATS landscape. Progress in Physical Geography, 36(3), 400-420. doi:10.1177/0309133312439594

[114]. Lambin EF, Tran A, Vanwambeke SO, Linard C, Soti V. (2010). Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. Int J Health Geography, 9(54), doi: 10.1186/1476-072X-9-54.

[115]. Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J. Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of U.S. adults.JAMA. 1998;279:1703–1708

[116]. Lantz PM, House JS, Mero RP, Williams DR. Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. J. Health Soc. Behav. 2005;46:274–288.

[117]. Lathey, V., Guhathakurta, S., & Aggarwal, R. (2009). The impact of subregional variations in urban sprawl on the prevalence of obesity and related morbidity. Journal of Planning Education and Research, 29(2), 127-141. doi:10.1177/0739456X09348615.

[118]. Layte R, Whelan CT. Explaining social class inequalities in smoking: the role of education, self-efficacy, and deprivation. Eur. Sociol. Rev. 2009;25:399–410.

[119]. Lee WJ, Blair A, Hoppin JA, Lubin JH, Rusiecki JA, Sandler DP, Dosemeci M, Alavanja MC. (2004). Cancer incidence among pesticide applicators exposed to chlorpyrifos in the Agricultural Health Study. J Natl Cancer Inst., 96(23):1781-1789.

[120]. Leitao, A. Botequila; Miller, J.; Ahern, J., McGarigal, K. (2006). Measuring Landscapes: A Planner's Handbook. Island Press, Washington DC.

[121]. Lejano, R. P., Ballesteros, F., & Tallod, R. (2012). Patchwork of land use, tapestry of risk. Journal Of Environmental Planning & Management, 55(1), 1-15. doi:10.1080/09640568.2011.574876.

[122]. Lerro CC, Koutros S, Andreotti G, Friesen MC, Alavanja MC, Blair A, Hoppin JA, Sandler DP, Lubin JH, Ma X, Zhang Y, Beane Freeman LE. (2015). Organophosphate insecticide use and cancer incidence among spouses of pesticide applicators in the Agricultural Health Study. Occup Environ Med., 72(10):736-44. doi: 10.1136/oemed-2014-102798.

[123]. Li X1, Sundquist K, Sundquist J. (2012a). Neighborhood deprivation and prostate cancer mortality: a multilevel analysis from Sweden. 2012. Prostate Cancer and Prostatic Diseases 15, 128-134.

[124]. Li, C., Li, J., & Wu, J. (2013). Quantifying the speed, growth modes, and landscape pattern changes of urbanization: A hierarchical patch dynamics approach. Landscape Ecology, 28(10), 1875-1888. doi:10.1007/s10980-013-9933-6

[125]. Li, Xinjun, Sundquist, Jan, Calling, Susanna, Zöller, Bengt and Sundquist, Kristina. (2012b). Gynecologic oncology 127(2). p.283-289.

[126]. Link, B. G., & Phelan, J. C. (1996). Understanding sociodemographic differences in health--the role of fundamental social causes. American Journal of Public Health, 86(4), 471–473.

[127]. Liu, J., Dietz, T., Carpenter, S. R., Folke, C., Alberti, M., Redman, C. L., and & ... Provencher, W. (2007). Coupled Human and Natural Systems. Ambio, (8), 639. doi:10.2307/25547831

[128]. Lopez, R. P., & Hynes, H. P. (2006). Obesity, physical activity, and the urban environment: public health research needs. doi: 10.1186/1476-069X-5-25

[129]. Lopez-Zetina, J., Lee, H., and Friis, R. (2006). The link between obesity and the built environment: Evidence from an ecological analysis of obesity and vehicle miles of travel in California. Health and Place, 12, p.656-664.

[130]. Luck, M. M., & Wu, J. J. (2002). A gradient analysis of urban landscape pattern: A case study from the Phoenix metropolitan region, Arizona, USA. Landscape Ecology, 17(4), 327-339. doi:10.1023/A:1020512723753.

[131]. Mackenbach, J. D., Rutter, H., Compernolle, S., Glonti, K., Oppert, J., Charreire, H., & ... Lakerveld, J. (2014). Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. BMC Public Health, 14(1), 1-28. doi:10.1186/1471-2458-14-233

[132]. Mai Stafford and Michael Marmot. (2003). Neighbourhood deprivation and health: does it affect us all equally? Int. J. Epidemiol. (2003) 32 (3):357-366.doi: 10.1093/ije/dyg084

[133]. Mannino, D. (2002). COPD: Epidemiology, prevalence, morbidity and mortality, and disease heterogeneity. Chest, 121(5), S121-126S.

[134]. McCann, B.A. and Ewing, R. (2003). Measuring the Health Effects of Sprawl: A National Analysis of Physical Activity, Obesity and Chronic Disease. Smart Growth America Surface Transportation Policy Project.

[135]. McGarigal, K. (2014). FRAGSTATS HELP. http://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf

[136]. McGarigal, K., Marks, B. (1995), FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure, General Technical Report PNW-GTR-351, USDA Forest Service.

[137]. McHugh, M., Kachroo, K., Liu, S., D’Amelio, M., Dong, A., Hong, M., . . . Etzel, K. (2010). Assessing environmental and occupational risk factors for lung cancer in Mexican–Americans. Cancer Causes & Control, 21(12), 2157-2164.

[138]. McKenzie, J.M., Robert R. Pinger, R.R., and Kotecki, J.E. (2012). An Introduction to Community Health, Seventh Edition, Jones & Bartlett Learning.

[139]. Mehta, N. K., House, J. S., & Elliott, M. R. (2015). Dynamics of health behaviours and socioeconomic differences in mortality in the USA. Journal of

Epidemiology & Community Health, 69(5), 416-422. doi:10.1136/jech-2014-204248

[140]. Messer, L., Laraia, B., Kaufman, J., Eyster, J., Holzman, C., Culhane, J., & ... O'Campo, P. (2006). The development of a standardized neighborhood deprivation index. Journal of Urban Health: Bulletin of the New York Academy of Medicine, 83(6), 1041-1062.

[141]. Midekisa, A., Senay, G., Henebry, G., Semuniguse, P., & Wimberly, M. (2012). Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. Malaria Journal, 11(1), 165.

[142]. Moore, P. (2011). A model to embed health outcomes into land-use planning. Community Development, 42(4), 525-540. doi: 10.1080/15575330.2011.593265

[143]. Moraga, P., & Montes, F. (2011). Detection of spatial disease clusters with LISA functions. Statistics In Medicine, 30(10), 1057-1071.

[144]. Nandi, A, MM Glymour, and SV Subramanian. n.d. "Association Among Socioeconomic Status, Health Behaviors, and All-Cause Mortality in the United States." Epidemiology25, no. 2: 170-177. Science Citation Index, EBSCOhost (accessed September 27, 2015).

[145]. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 06/04/2014.

[146]. Nordgren, T.M. and Bailey, K.L. (2016). Pulmonary health effects of agriculture. Current Opinion in Pulmonary Medicine, 22(2), p.144-149.

[147]. Northridge, M.E., Sclar, E.D., Biswas, P. (2003). Sorting out the connections between the built environment and health: A conceptual framework for navigating pathways and planning healthy cities. Journal of Urban Health, 80 (4) , pp 556-568

[148]. Orenstein, D., Frenkel, A., & Jahshan, F. (n.d.). Methodology Matters: Measuring Urban Spatial Development Using Alternative Methods. 41(1), 3-23.

[149]. Pampel, F. C., Krueger, P. M., & Denney, J. T. (2010). Socioeconomic Disparities in Health Behaviors. Annual Review of Sociology, 36, 349–370. http://doi.org/10.1146/annurev.soc.012809.102529

[150]. Park, S. 2013. "Spatiotemporal Landscape Pattern Change in Response to Future Urbanization in Maricopa County, Arizona, USA." Landscape Research 38, no. 5: 625-648. Business Source Complete, EBSCOhost (accessed July 1, 2014).

[151]. Patz, J. A., Daszak, P., Tabor, G. M., Aguirre, A., Pearl, M., Epstein, J., & ... Members of the Working Group on Land Use Change and Disease, E. (2004). Unhealthy Landscapes: Policy Recommendations on Land Use Change and Infectious Disease Emergence. Environmental Health Perspectives, (10). 1092.

[152]. Paudel, S., & Yuan, F. (2012). Assessing landscape changes and dynamics using patch analysis and GIS modeling. International Journal of Applied Earth Observations & Geoinformation, 1666-76. doi:10.1016/j.jag.2011.12.003

[153]. Peng, J., Wang, Y., Zhang, Y., Wu, J., Li, W., & Li, Y. (2010). Evaluating the effectiveness of landscape metrics in quantifying spatial patterns. Ecological Indicators, 10217-223. doi:10.1016/j.ecolind.2009.04.017

[154]. Plantinga, Andrew J., and Bernell Stephanie. 2007. "Can Urban Planning Reduce Obesity? The Role of Self-Selection in Explaining the Link between Weight and Urban Sprawl." Review of Agricultural Economics, 2007. 557. JSTOR Journals, EBSCOhost (accessed May 31, 2015).

[155]. Prescotta, E., Vestbob, J. (1999). Socioeconomic status and chronic obstructive pulmonary disease. Thorax, 54:737-741 doi:10.1136/thx.54.8.737.

[156]. Raviv S., Hawkins, K.A., DeCamp Jr., M.M., Kalhan,R. (2011). Lung Cancer in Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine, 183(9), pp. 1138-1146. doi: 10.1164/rccm.201008-1274CI

[157]. Richardson, E.)., Pearce, J.)., Shortt, N.)., & Mitchell, R.). (2013). A Regional Measure of Neighborhood Multiple Environmental Deprivation: Relationships with Health and Health Inequalities.Professional Geographer, 65(1), 153-170. doi:10.1080/00330124.2012.660457

[158]. Ritz, B., & Rull, R. (2008). Assessment of environmental exposures from agricultural pesticides in childhood leukaemia studies: Challenges and opportunities. Radiation Protection Dosimetry, 132(2), 148-155.

[159]. Ritz, B., & Rull, R. (2008). Assessment of environmental exposures from agricultural pesticides in childhood leukemia studies: Challenges and opportunities. Radiation Protection Dosimetry, 132(2), 148-155.

[160]. Rogers RG, Hummer RA, Nam CB. Living and Dying in the USA: Behavioral, Health, and Social Differentials of Adult Mortality. San Diego, CA: Academic; 2000.

[161]. Ross, C. L., Orenstein, M., & Botchwey, N. (2014). Health Impact Assessment in the United States. New York, NY, US: Springer. Retrieved from http://www.ebrary.com.prx.library.gatech.edu.

[162]. Rull, R. P., & Ritz, B. (2003). Historical pesticide exposure in California using pesticide use reports and land-use surveys: an assessment of misclassification error and bias. Environmental Health Perspectives, 111(13), 1582–1589.

[163]. Rushton,G. . Armstrong, M. P, Gittler J., . Greene B. R, Pavlik, C. E., West M. M., Zimmerman D. L. (2008). Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice. CRC Press, Taylor and Francis Group, Boca Raton, FL.

[164]. Sabanayagam, C., & Shankar, A. (2012). Income is a stronger predictor of mortality than education in a national sample of US adults. Journal of Health, Population, And Nutrition, 30(1), 82-86.

[165]. Sallis, J., Floyd, M., Rodriguez, D., & Saelens, B. (n.d). Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease. Circulation, 125(5), 729-737.

[166]. Samimi, A., & Mohammadian, A. K. (2010). Health Impacts of Urban Development and Transportation Systems. Journal of Urban Planning & Development, 136(3), 208-213. doi: 10.1061/(ASCE)UP.1943-5444.0000020

[167]. Saxton-Ross, A. (2009). Physical activity and the urban form: How the built environment, residential density, residential segregation, income inequality, race, gender and education, influence physical activity in Atlanta, Georgia. Dissertation Abstracts International Section A, 70,

[168]. Schenker, M. (1998). Respiratory Health Hazards in Agriculture. Conference Proceedings of the American Thoracic Society. American Journal of Respiratory and Critical Care Medicine, 158(5). https://www.thoracic.org/statements/resources/archive/agriculture1-79.pdf

[169]. Schikowski T, Sugiri D, Ranft U et al. Long-term air pollution exposure and living close to busy roads are associated with COPD in women. Respir. Res. 2005; 6: 152.

[170]. Schweitzer, L., & Zhou, J. (2010). Neighborhood air quality, respiratory health, and vulnerable populations in compact and sprawled regions. Journal of the American Planning Association, 76(3), 363-371.

[171]. Seto, K., & Fragkias, M. (2005). Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landscape Ecology, 20(7), 871-888. doi:10.1007/s10980-005-5238-8.

[172]. Siahpush M, Borland R, Taylor J, Singh GK, Ansari Z, Serraglio A. The association of smoking with perception of income inequality, relative material well-being, and social capital. Soc. Sci. Med.2006a;63:2801–2812. [PubMed]

[173]. Siahpush M, McNeil A, Hammond D, Fong GT. Socioeconomic and country variations in knowledge of health risks of tobacco smoking and toxic constituents of smoke: results from the 2002 International Tobacco Control (ITC) Four Country Survey. Tob. Control. 2006b;15.

[174]. Siegel, R. L., Miller, K. D. and Jemal, A. (2015), Cancer statistics, 2015. CA: A Cancer Journal for Clinicians, 65: 5–29. doi:10.3322/caac.21254

[175]. Signorello, L. B., Cohen, S. S., Williams, D. R., Munro, H. M., Hargreaves, M. K., & Blot, W. J. (2014). Socioeconomic Status, Race, and Mortality: A Prospective Cohort Study. American Journal of Public Health, 104(12), e98-e107.

[176]. Singh, G. K. (2003). Area Deprivation and Widening Inequalities in US Mortality, 1969–1998. American Journal of Public Health, 93(7), 1137–1143.

[177]. Sloggett A, Joshi H.(1994). Higher mortality in deprived areas: community or personal disadvantage? BMJ.1994;309:1470–1474.

[178]. Stafford, M. M., & Marmot, M. M. (2003). Neighbourhood deprivation and health: Does it affect us all equally? International Journal of Epidemiology, 32(3), 357-366.

[179]. Stafford, M., Gimeno, D., & Marmot, M. G. (2008). Neighbourhood characteristics and trajectories of health functioning: a multilevel prospective analysis. European Journal of Public Health, 18(6), 604-610.

[180]. Steiner, F. (2004). Urban human ecology. Urban Ecosystems, 7(3), 179-197.

[181]. Steiner, F. (2011). Landscape ecological urbanism: Origins and trajectories. Landscape & Urban Planning, 100(4), 333. doi:10.1016/j.landurbplan.2011.01.020

[182]. Steiner, F., Simmons, M., Gallagher, M., Ranganathan, J., & Robertson, C. (2013). The ecological imperative for environmental design and planning. Frontiers in Ecology & the Environment, 11(7), 355-361. doi:10.1890/130052.

[183]. Stimpson J, Ju H, Raji M, Eschbach K. Neighborhood deprivation and health risk behaviors in NHANES III... Third National Health and Nutrition Examination Survey. American Journal Of Health Behavior [serial online]. March 2007;31(2):215-222. Available from: CINAHL, Ipswich, MA. Accessed June 2, 2015.

[184]. Stockard, J., Doyle, S., Schlossberg, M., & Kelly-Schwartz, A. C. (2004). Is sprawl unhealthy? A multilevel analysis of the relationship of metropolitan sprawl to the health of individuals. Journal of Planning Education and Research, 24(2), 184-196. doi:10.1177/0739456X04267713

[185]. Su, S. S., Xiao, R. R., Jiang, Z. Z., & Zhang, Y. Y. (2012). Characterizing landscape pattern and ecosystem service value changes for urbanization impacts

at an eco-regional scale. Applied Geography, 34(2), 295-305. doi:10.1016/j.apgeog.2011.12.001

[186]. Sundquist, J., Li, X. and Sundquist, a. K. (2012), Neighborhood Deprivation and Mortality in Women with Breast Cancer: A Multilevel Analysis From Sweden. The Breast Journal, 18: 625–627. doi: 10.1111/tbj.12035

[187]. Sundquist, K.)., Eriksson, U.)., Ohlsson, H.)., & Mezuk, B.). (2015). Neighborhood walkability, deprivation and incidence of type 2 diabetes: A population-based study on 512,061 Swedish adults. Health And Place, 3124-30. doi:10.1016/j.healthplace.2014.10.011

[188]. Testi, E. Ivaldi, A. Busi. (2004). An Index Of Material Deprivation for Geographical Areas.

[189]. Townshend, T., & Lake, A. A. (2009). Obesogenic urban form: Theory, policy and practice. Health & Place, 15(4), 909-916. doi: 10.1016/j.healthplace.2008.12.002

[190]. US Census Bureau .(2010). Poverty thresholds by Size of Family and Number of Children. https://www.census.gov/hhes/www/poverty/data/threshld/index.html.

[191]. Uuemaa, E. E., Mander, Ü. Ü., & Marja, R. R. (2013). Trends in the use of landscape spatial metrics as landscape indicators: A review. Ecological Indicators, 28100-106. doi:10.1016/j.ecolind.2012.07.018

[192]. Uuemaa, E., Antrop, M., Roosaare, J., Marja, R., & Mander, Ü. (2009). Landscape Metrics and Indices: An Overview of Their Use in Landscape Research. Living Reviews Inin Landscape Research, 35-28.

[193]. Vargo, Jason, Habeeb, Dana, & Stone, Brian. (2013). The importance of land cover change across urban–rural typologies for climate modeling. Journal of Environmental Management, 114, 243-252.

[194]. Vazquez-Prokopec, G., Spillmann, C., Zaidenberg, M., Gürtler, R., Kitron, U., & Kittayapong, P. (2012). Spatial Heterogeneity and Risk Maps of Community Infestation by Triatoma infestans in Rural Northwestern Argentina (T. infestans Infestation Risk). 6(8), E1788.

[195]. Von Maltzahn, K.E. (1994). “Nature as Landscape” in Nature as Landscape: Dwelling and Understanding. Montréal : McGill-Queen's University Press, c1994.

[196]. Waller, A. and Gotway, C. (2004). “Spatial clustering of Health Events: Regional Count Data” in Applied Spatial Statistics for Public Health Data. John Wiley and Sons, pg.200-271.

[197]. Weichenthal, S., Moase, C., & Chan, P. (2010). A Review of Pesticide Exposure and Cancer Incidence in the Agricultural Health Study Cohort.Environmental Health Perspectives, 118(8), 1117–1125. http://doi.org/10.1289/ehp.0901731

[198]. Weichenthal, S., Moase, C., & Chan, P. (2010). A Review of Pesticide Exposure and Cancer Incidence in the Agricultural Health Study Cohort. Environmental Health Perspectives, 118(8), 1117–1125. http://doi.org/10.1289/ehp.0901731

[199]. Weinhold, B. (2007). Occupational Health: Agriculture–COPD Link Bolstered.Environmental Health Perspectives, 115(9), A444.

[200]. Wells, N. M., Evans, G. W., & Yang, Y. Z. (2010). Environments Andand Health: Planning Decisions Asas Public-Health Decisions. Journal Ofof Architectural Andand Planning Research, 27(2), 124-143.

[201]. Wild C. 2005. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiology Biomarkers and Prevention.;14:1847–1850. DOI.org/10:1158/10559965.

[202]. Williams DR and Collins C. (2001). Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Reports, 116, 404-416.

[203]. Williams DR, Neighbors HW, Jackson JS. (2003) Racial/ethnic discrimination and health: Findings from community studies. American Journal of Public Health, 93, 200-208.

[204]. Wu, J. (2013). Landscape sustainability science: ecosystem services and human well-being in changing landscapes. Landscape Ecology, 28(6), 999-1023. doi:10.1007/s10980-013-9894-9.

[205]. Yang, W., Dietz, T., Liu, W., Luo, J., & Liu, J. (2013). Going Beyond the Millennium Ecosystem Assessment: An Index System of Human Dependence on Ecosystem Services. Plos ONE, 8(5), 1-9. doi:10.1371/journal.pone.0064581

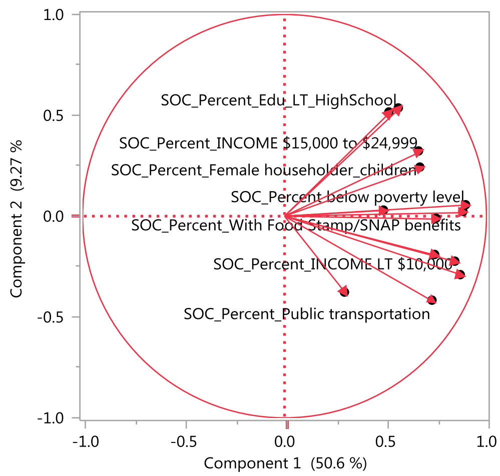
[206]. Yarrow, M., & Salthe, S. (2008). Ecological boundaries in the context of hierarchy theory. Biosystems, 92(3), 233-244. doi:10.1016/j.biosystems.2008.03.001

[207]. Zhao, Z., & Kaestner, R.. (2010). Effects of urban sprawl on obesity. Journal of Health Economics, 29(6), 779-787. doi:10.1016/j.jhealeco.2010.07.006

[208]. Zhu, H., D.L. Reichard, R.D. Fox, R.D. Brazee and H.E. Ozkan. (1994). Simulation of drift of discrete sizes of water droplets from field sprayers. Transactions of the ASAE 37(5):1401-1407.

[209]. Zonneveld, I. (1994). Landscape Ecology and ecological networks in Landscape Ecological Planning and Ecological Networks. Cook, E.A. and Van Lier (eds.). Elsevier, Amsterdam, Netherlands.

APPENDIX A: NEIGHBORHOOD DERIVATION INDEX CONSTRUCTION



Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	6.5823	50.633		50.633
2	1.2046	9.266		59.899
3	0.9761	7.509		67.407
4	0.8276	6.366		73.774
5	0.7521	5.786		79.559
6	0.6062	4.663		84.222
7	0.5225	4.019		88.242
8	0.3988	3.068		91.310
9	0.3856	2.966		94.276
10	0.3362	2.586		96.862
11	0.1938	1.491		98.353
12	0.1237	0.952		99.304
13	0.0904	0.696		100.000

Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin11	Prin12	Prin13
SOC_Percent_Edu_LT_HighSchool	0.56994	0.53730	0.29970	0.06449	-0.35262	-0.18516	0.03815	-0.13560	-0.12243	0.31179	0.01839	0.03098	0.02308
SOC_Percent_Female householder_children	0.67294	0.24052	-0.25492	0.07887	0.38685	0.20242	-0.41651	-0.02620	0.04094	0.18090	0.12774	0.04261	-0.02770
SOC_Percent below poverty level	0.90229	0.05459	0.05389	-0.05288	-0.21365	0.08757	-0.03013	-0.06036	-0.06111	-0.16451	0.15398	-0.22159	-0.12533
SOC_Percent_Vacant housing units	0.75017	-0.19684	0.01986	-0.11020	0.24832	-0.04128	0.23469	-0.47576	0.20168	0.00910	-0.01021	0.01327	0.01154
SOC_Percent_No vehicles available	0.87595	-0.28975	-0.02109	-0.19049	-0.08343	-0.09790	0.05620	0.11047	-0.08376	0.01331	-0.02111	0.19683	-0.18367
SOC_Percent_COSTS GT 35.0 percent of HH incom	0.30809	-0.38127	0.67047	0.49332	0.17822	0.16500	-0.02822	0.03692	-0.06829	0.02961	-0.00830	0.00067	-0.00858
SOC_Percent_Public transportation	0.73235	-0.42089	-0.01865	-0.27530	0.07846	-0.09993	0.11787	0.26497	0.07525	0.28606	0.05578	-0.11439	0.08027
SOC_Percent_INCOME LT $10,000	0.84586	-0.22604	-0.11702	-0.07136	-0.08961	0.03124	-0.08460	-0.07752	-0.34994	-0.17514	0.08352	0.07228	0.16474
SOC_Percent_INCOME $10,000 to $14,999	0.75352	-0.01089	0.12283	0.12653	-0.20584	-0.32542	-0.28087	0.08137	0.36230	-0.17688	0.00176	0.03884	0.05615
SOC_Percent_INCOME $15,000 to $24,999	0.66754	0.31440	0.11824	-0.14104	-0.14316	0.53984	0.20641	0.13945	0.18804	-0.06726	-0.01691	0.07026	0.05081
SOC_Percent_INCOME $25,000 to $34,999	0.52071	0.51048	0.19608	-0.09131	0.50505	-0.26470	0.17328	0.16380	-0.10368	-0.16837	-0.01351	-0.01330	0.00521
SOC_Percent_With cash public assistance income	0.49598	0.03139	-0.50108	0.62262	-0.04480	-0.06448	0.31212	0.08107	0.02171	0.01331	0.05844	0.00728	-0.00131
SOC_Percent_With Food Stamp/SNAP benefits	0.88653	0.02026	-0.18421	0.06197	-0.00727	0.06498	-0.13175	-0.01155	-0.08339	0.02175	-0.37242	-0.08932	-0.00336

Socioeconomic Index is the first Principal Component, with the following equation

SOC_PCA_INDEX = (0.0397087014102446 * :SOC_Percent_Edu_LT_HighSchool) + (0.0475642459130662 * :SOC_Percent_Female householder_children) + (0.0288047200735408 * :SOC_Percent below poverty level) + (0.0364723756690012 * :SOC_Percent_Vacant housing units) + (0.0342574332571883 * :SOC_Percent_No vehicles available) + (0.0098556627755686 * :SOC_Percent_COSTS GT 35.0 percent of HH income) + (0.0361666411532527 * :SOC_Percent_Public transportation) + (0.0474600369464659 * :SOC_Percent_INCOME LT $10,000) + (0.0732055119265184 *:SOC_Percent_INCOME $10,000 to $14,999) + (0.0398052312698556 * :SOC_Percent_INCOME $15,000 to $24,999) + (0.0374637241975227 * :SOC_Percent_INCOME $25,000 to $34,999) + (0.119779381362708 * :SOC_Percent_With cash public assistance income) + (0.0392144438512313 * :SOC_Percent_With Food Stamp/SNAP benefits) + (-3.99369979639954)

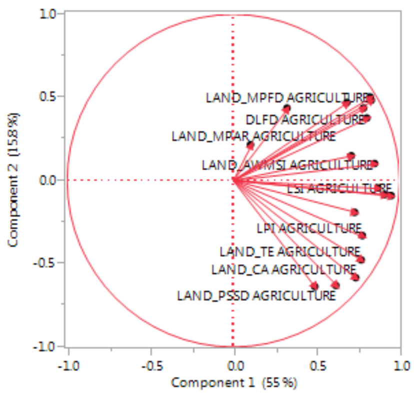
Figure 47. Loading matrix and equation used to calculate the Neighborhood Deprivation Index

APPENDIX B: LAND USE INDICES

Level 1 PCA Index - AGRICULTURE

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	9.8936	54.964		54.964
2	2.8374	15.763		70.728
3	1.3173	7.318		78.046
4	1.1675	6.486		84.532
5	0.7972	4.429		88.961
6	0.6265	3.480		92.442
7	0.4236	2.353		94.795
8	0.3256	1.809		96.604
9	0.1468	0.815		97.419
10	0.1264	0.702		98.121
11	0.0991	0.551		98.672
12	0.0728	0.404		99.076
13	0.0582	0.323		99.400
14	0.0500	0.278		99.677
15	0.0374	0.208		99.885
16	0.0164	0.091		99.976
17	0.0034	0.019		99.995
18	0.0008	0.005		100.000



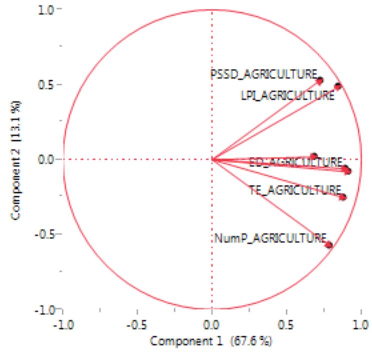
Loading Matrix

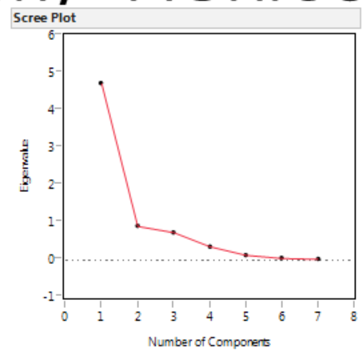
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS AGRICULTURE	0.71403	0.14870	-0.17563	-0.16980	-0.17975	0.38876	0.45772	-0.08143	-0.02411
MPI AGRICULTURE	0.49269	-0.64012	0.39953	-0.28864	0.09915	-0.04714	0.00075	0.18437	0.18720
MNN AGRICULTURE	0.32859	0.43436	-0.04626	-0.27339	0.77106	0.04367	0.04595	-0.15549	0.04482
IJI AGRICULTURE	0.78336	0.43038	-0.15798	-0.19905	-0.02714	0.13700	-0.15404	0.16134	0.04708
LAND_NumP AGRICULTURE	0.73456	-0.19037	-0.38585	0.47044	0.10677	-0.07748	-0.00028	0.02089	0.04803
LAND_PSCoV AGRICULTURE	0.91031	-0.08425	-0.10964	0.03386	0.07996	-0.00265	-0.25943	-0.11176	-0.17685
LAND_PSSD AGRICULTURE	0.62215	-0.63021	0.32572	-0.24221	0.08442	-0.00182	0.04440	0.05744	-0.10634
LAND_TE AGRICULTURE	0.77457	-0.47970	-0.20161	0.31276	0.10694	-0.06260	0.09437	0.00488	0.00360
LAND_ED AGRICULTURE	0.87284	-0.04745	-0.18609	-0.01907	-0.19576	0.18377	-0.22149	-0.16632	0.17613
LAND_MSI AGRICULTURE	0.68853	0.46912	0.40001	0.19642	-0.06600	-0.18125	0.00456	-0.13656	-0.04279
LAND_AWMSI AGRICULTURE	0.85195	0.09749	0.32824	0.02913	-0.12404	-0.26349	0.11690	-0.14530	0.03180
LAND_MPFD AGRICULTURE	0.82592	0.49792	0.10644	-0.02465	-0.06972	-0.17480	0.03604	0.05770	0.01863
LAND_AWMPFD AGRICULTURE	0.83432	0.47796	0.10151	-0.02309	-0.07816	-0.17196	0.06509	0.06564	0.03795
LPI AGRICULTURE	0.77685	-0.33621	0.18735	-0.32570	-0.13413	0.19895	-0.15043	-0.15621	-0.07562
LSI AGRICULTURE	0.95381	-0.09265	-0.21402	0.09666	-0.01472	0.03685	-0.02068	0.01918	0.06066
LAND_CA AGRICULTURE	0.74249	-0.58989	-0.06347	0.19240	0.13034	-0.07018	0.12155	0.03541	-0.05566
DLFD AGRICULTURE	0.80480	0.37187	-0.15000	-0.13557	0.01496	0.02900	-0.03402	0.33140	-0.11752
LAND_MPAR AGRICULTURE	0.10902	0.21201	0.60053	0.59190	0.13781	0.44680	-0.05260	0.08709	0.00766

	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS AGRICULTURE	0.05048	0.06376	0.01934	-0.01439	0.03122	-0.00249	0.00614	0.00014	-0.00003
MPI AGRICULTURE	0.03925	0.12860	0.05366	-0.01536	0.02661	-0.03960	0.00339	-0.00012	0.00027
MNN AGRICULTURE	-0.03919	-0.00855	-0.00562	0.00917	-0.00346	0.00215	0.00104	-0.00017	-0.00006
IJI AGRICULTURE	0.21420	-0.06455	-0.10959	0.03616	-0.01846	-0.03619	-0.00693	0.00119	-0.00034
LAND_NumP AGRICULTURE	0.04993	0.04392	0.05525	0.14653	0.05578	0.04086	0.00312	-0.01021	-0.00103
LAND_PSCoV AGRICULTURE	0.09495	0.05720	0.07028	-0.09400	0.08141	-0.02921	0.00705	0.00104	0.00143
LAND_PSSD AGRICULTURE	0.03874	-0.06266	-0.06054	0.00854	0.03096	0.13081	-0.00633	0.00493	0.00043
LAND_TE AGRICULTURE	-0.01919	-0.01847	-0.03117	-0.01268	-0.04191	-0.03807	-0.00755	0.04526	0.00245
LAND_ED AGRICULTURE	-0.10360	0.00360	-0.05903	-0.04105	0.01593	0.04996	0.04889	0.00213	0.00061
LAND_MSI AGRICULTURE	0.03194	0.18850	-0.08449	0.02014	-0.07007	0.02261	-0.00548	-0.00078	0.00240
LAND_AWMSI AGRICULTURE	-0.05318	-0.10016	-0.06705	0.02310	0.11334	-0.06653	-0.01710	-0.00485	-0.00321
LAND_MPFD AGRICULTURE	0.02324	-0.04485	0.10219	-0.02644	-0.04140	0.03384	0.02394	0.00959	-0.01958
LAND_AWMPFD AGRICULTURE	0.00731	-0.08633	0.09810	-0.01469	-0.02220	0.02220	0.01437	-0.00088	0.02059
LPI AGRICULTURE	-0.06080	-0.03273	0.07396	0.11877	-0.07501	-0.04307	-0.01189	0.00035	-0.00025
LSI AGRICULTURE	-0.04992	-0.00827	0.01272	-0.07549	-0.04085	0.02287	-0.09791	-0.01609	-0.00181
LAND_CA AGRICULTURE	0.00407	-0.04972	-0.04596	-0.04294	-0.08231	-0.04156	0.05306	-0.02972	-0.00126
DLFD AGRICULTURE	-0.19982	0.06254	-0.03996	0.02243	0.03171	-0.01514	0.00899	0.00033	-0.00011
LAND_MPAR AGRICULTURE	-0.01403	-0.04763	0.01796	-0.01016	0.01496	-0.00610	0.00013	-0.00005	-0.00049

Figure 48. Eigen values and Loading Matrix for Level 1 Agriculture Index

Level 2 PCA (GEOM) - AGRICULTURE





Eigenvalues

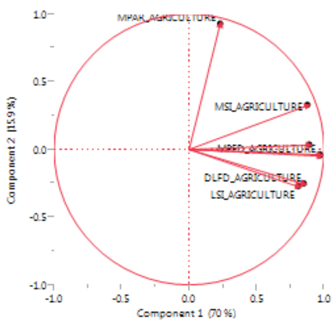
Number	Eigenvalue	Percent	20 40 60 80	Cum Percent	ChiSquare	DF	Prob>ChiSq
1	4.7309	67.584		67.584	7588.52	17.313	<.0001*
2	0.9155	13.079		80.663	3529.16	18.281	<.0001*
3	0.7487	10.695		91.358	2781.40	13.391	<.0001*
4	0.3721	5.316		96.674	1538.55	8.880	<.0001*
5	0.1416	2.023		98.697	496.567	5.042	<.0001*
6	0.0569	0.813		99.510	60.005	2.054	<.0001*
7	0.0343	0.490		100.000	0.000	.	.

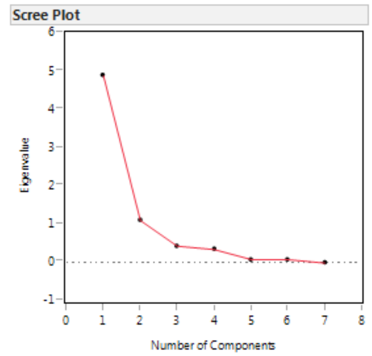
Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
MedPS_AGRICULTURE	0.68521	0.01738	0.63347	0.35339	0.06303	0.00650	0.00100
PSSD_AGRICULTURE	0.72342	0.52825	-0.38180	0.19485	0.02339	0.09511	0.06548
PSCoV_AGRICULTURE	0.91476	-0.08165	-0.00746	-0.26869	0.28746	0.02978	-0.02783
NumP_AGRICULTURE	0.78705	-0.57212	-0.17700	0.06178	-0.01544	-0.07369	0.11146
LPI_AGRICULTURE	0.84443	0.48388	0.04851	-0.13022	-0.06533	-0.17065	-0.01017
TE_AGRICULTURE	0.87864	-0.25344	-0.31998	0.19660	-0.07854	0.00258	-0.12860
ED_AGRICULTURE	0.89261	-0.06162	0.25594	-0.27864	-0.20921	0.11127	0.01262

Figure 49. Eigen values and Loading Matrix for Level 2 Agriculture Geometry Index

Level 2 PCA (SHAPE) - AGRICULTURE





Eigenvalues

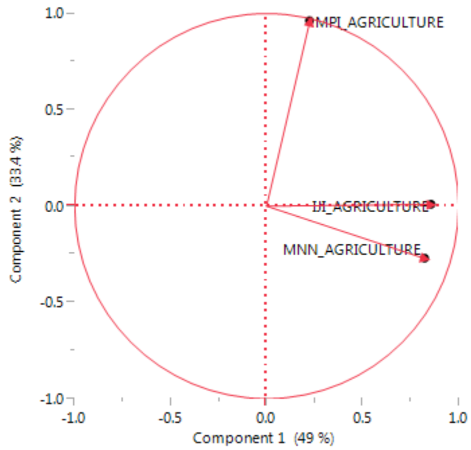
Number	Eigenvalue	Percent	Cum Percent	ChiSquare	DF	Prob>ChiSq
1	4.9010	70.014	70.014	10845.0	17.426	<.0001*
2	1.1154	15.934	85.948	6374.08	18.865	<.0001*
3	0.4423	6.318	92.266	4742.85	13.426	<.0001*
4	0.3541	5.059	97.325	4090.72	8.712	<.0001*
5	0.0945	1.350	98.675	2802.10	4.846	<.0001*
6	0.0913	1.304	99.979	2630.93	1.866	<.0001*
7	0.0015	0.021	100.000	0.000	.	.

Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
MPAR_AGRICULTURE	0.23122	0.92934	0.28206	0.00882	0.05685	0.00283	0.00074
LSI_AGRICULTURE	0.81247	-0.27571	0.36093	0.34345	-0.07305	-0.10154	-0.00047
MSI_AGRICULTURE	0.87740	0.32393	-0.26293	-0.03509	-0.23421	-0.00088	-0.00350
MPFD_AGRICULTURE	0.97102	-0.03861	-0.09648	-0.16449	0.06938	-0.11711	0.02717
AWMSI_AGRICULTURE	0.89541	0.03221	-0.23299	0.32403	0.10803	0.16201	0.00277
AWMPFD_AGRICULTURE	0.97294	-0.04911	-0.08661	-0.13758	0.11677	-0.10107	-0.02636
DLFD_AGRICULTURE	0.85223	-0.25657	0.30368	-0.28966	-0.03053	0.17555	0.00008

Figure 50. Eigen values and Loading Matrix for Level 2 Agriculture Shape Index

Level 2 PCA (INT) - AGRICULTURE



Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent	ChiSquare	DF	Prob>ChiSq
1	1.4708	49.026		49.026	239.974	2.615	<.0001*
2	1.0025	33.418		82.444	96.594	1.527	<.0001*
3	0.5267	17.556		100.000	0.000	.	.

Loading Matrix

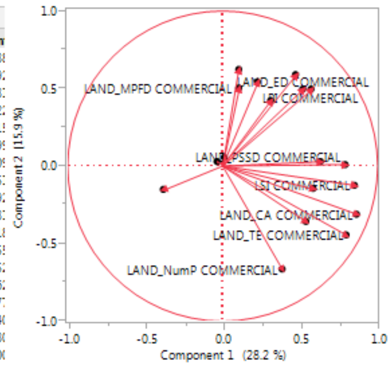
	Prin1	Prin2	Prin3
MPI_AGRICULTURE	0.22623	0.96332	0.14435
IJI_AGRICULTURE	0.85867	0.00871	-0.51246
MNN_AGRICULTURE	0.82601	-0.27289	0.49319

Figure 51. Eigen values and Loading Matrix for Level 2 Agriculture Interspersion Index

PCA Index - COMMERCIAL

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percen
1	5.0738	28.188		28.18
2	2.8627	15.904		44.09
3	2.4101	13.389		57.48
4	1.0875	6.041		63.52
5	1.0246	5.692		69.21
6	0.9871	5.484		74.69
7	0.9197	5.110		79.80
8	0.8537	4.743		84.55
9	0.6012	3.340		87.89
10	0.5201	2.890		90.78
11	0.4927	2.737		93.51
12	0.4386	2.437		95.95
13	0.2695	1.497		97.45
14	0.2177	1.209		98.66
15	0.1109	0.616		99.27
16	0.0652	0.362		99.64
17	0.0432	0.240		99.88
18	0.0216	0.120		100.00



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS COMMERCIAL	0.31024	0.41880	-0.49403	0.14466	-0.00851	0.03491	0.03350	0.56308	0.24202
MPI COMMERCIAL	0.63395	0.02812	-0.17215	-0.26063	-0.23107	0.18218	-0.19669	-0.09879	-0.39091
MNN COMMERCIAL	-0.37770	-0.15505	0.14659	0.04103	-0.55990	0.44651	0.32504	0.27052	-0.04072
IJI COMMERCIAL	0.58468	-0.14874	0.05849	0.35571	0.15921	0.01783	0.49023	-0.07742	0.05605
LAND_NumP COMMERCIAL	0.38684	-0.66808	0.40615	0.12954	0.16554	-0.17464	-0.05226	0.16985	0.01269
LAND_PSCoV COMMERCIAL	0.53780	-0.35720	0.42781	-0.17034	-0.08859	0.13728	0.16441	-0.38018	0.24298
LAND_PSSD COMMERCIAL	0.79488	0.00979	-0.14342	-0.25044	-0.26782	0.23342	-0.04820	-0.11546	0.12330
LAND_TE COMMERCIAL	0.80219	-0.44658	0.15895	0.02743	0.06324	-0.09770	-0.16374	0.21925	0.00751
LAND_ED COMMERCIAL	0.51950	0.49341	-0.38755	0.10533	0.16200	-0.11628	0.20453	-0.25410	-0.13401
LAND_MSI COMMERCIAL	0.10402	0.61912	0.50579	-0.20637	0.08576	0.02952	-0.07748	0.00506	0.36934
LAND_AWMSI COMMERCIAL	0.47434	0.58643	0.23322	-0.12243	-0.03446	-0.13242	-0.31336	0.04680	-0.00040
LAND_MPFD COMMERCIAL	0.10852	0.50151	0.76587	0.12425	-0.09119	0.06213	0.14199	-0.01185	-0.01578
LAND_AWMPFD COMMERCIA	0.22762	0.53769	0.62786	0.21991	-0.08071	-0.06586	0.02077	0.15196	-0.32402
LPI COMMERCIAL	0.57339	0.49023	-0.50602	-0.00957	-0.03176	0.05832	0.20073	-0.07475	0.06921
LSI COMMERCIAL	0.85283	-0.12390	0.00883	0.19390	0.06286	-0.03460	0.13273	0.05585	-0.12263
LAND_CA COMMERCIAL	0.86623	-0.31066	-0.02029	-0.08642	-0.08648	0.05823	-0.15526	0.19653	0.05884
DLFD COMMERCIAL	0.01111	0.05096	-0.03450	0.58108	0.21669	0.62903	-0.42761	-0.16737	0.05124
LAND_MPAR COMMERCIAL	-0.03276	0.02882	0.13711	-0.49582	0.64214	0.44253	0.20873	0.21814	-0.15836

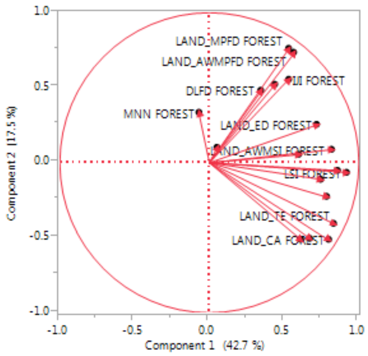
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS COMMERCIAL	-0.15176	-0.02225	-0.11702	0.13963	-0.14886	-0.05888	0.05456	0.00240	-0.00224
MPI COMMERCIAL	-0.38533	0.17447	0.12367	0.11327	-0.06967	0.00341	0.00175	0.00175	-0.00158
MNN COMMERCIAL	0.23540	0.14393	0.20130	0.00787	0.00583	-0.00467	0.00286	0.00121	0.00080
IJI COMMERCIAL	-0.26001	-0.24813	0.30418	0.04645	0.08331	0.00240	-0.01098	-0.00272	0.00379
LAND_NumP COMMERCIAL	0.05218	0.25400	-0.01697	0.18540	0.07303	0.12276	0.09276	0.00627	0.04119
LAND_PSCoV COMMERCIAL	0.08760	-0.02955	-0.14110	0.19745	-0.19775	-0.07454	-0.02612	0.03203	-0.00588
LAND_PSSD COMMERCIAL	0.02241	-0.24232	-0.11830	-0.10895	0.13175	0.07067	0.14137	-0.00317	-0.01685
LAND_TE COMMERCIAL	0.09813	0.07519	0.06343	-0.01045	0.06484	-0.06127	-0.04162	-0.04534	-0.10673
LAND_ED COMMERCIAL	0.24008	0.25647	0.06436	0.00243	0.05212	-0.15089	0.09029	0.00535	0.00604
LAND_MSI COMMERCIAL	-0.14865	0.26328	0.20779	-0.12366	0.02846	0.02699	0.00133	0.07205	-0.01336
LAND_AWMSI COMMERCIAL	0.27167	-0.22566	0.30152	0.12266	-0.07974	0.03487	-0.01104	-0.05704	0.03126
LAND_MPFD COMMERCIAL	-0.15447	0.09908	-0.21457	-0.04514	0.03095	-0.03426	0.00477	-0.13363	0.01509
LAND_AWMPFD COMMERCIA	0.05648	-0.16380	-0.16444	0.06303	0.05348	-0.00304	-0.00364	0.11297	-0.02177
LPI COMMERCIAL	0.10520	0.14061	-0.14898	0.11804	0.10909	0.16120	-0.12123	-0.00816	-0.00745
LSI COMMERCIAL	0.08291	0.05674	-0.01863	-0.29505	-0.26794	0.09611	-0.00641	0.00686	0.01078
LAND_CA COMMERCIAL	0.00600	-0.01446	-0.06360	-0.10870	0.14243	-0.12982	-0.08982	0.02713	0.07840
DLFD COMMERCIAL	0.03574	0.02575	0.00237	0.01803	0.00288	0.00351	-0.00130	0.00297	-0.00040
LAND_MPAR COMMERCIAL	0.07223	-0.05625	-0.02148	0.01098	-0.00044	0.00034	0.00176	-0.00890	0.00182

Figure 52. Eigen values and Loading Matrix for Level 1 Commercial Index

PCA Index - FORESTS

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	7.6938	42.743		42.743
2	3.1438	17.466		60.209
3	1.7054	9.475		69.684
4	1.2498	6.944		76.627
5	0.8806	4.892		81.519
6	0.7766	4.314		85.833
7	0.6833	3.796		89.630
8	0.6252	3.473		93.103
9	0.4031	2.239		95.342
10	0.2174	1.208		96.550
11	0.2041	1.134		97.684
12	0.1342	0.745		98.430
13	0.0923	0.513		98.943
14	0.0798	0.444		99.386
15	0.0674	0.374		99.760
16	0.0332	0.184		99.945
17	0.0054	0.030		99.975
18	0.0045	0.025		100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS FOREST	0.60010	0.05393	-0.33713	-0.05135	0.00660	0.12371	0.53873	0.44419	-0.02798
MPI FOREST	0.61445	-0.51400	0.33346	0.30462	-0.12680	0.07611	0.06311	-0.03662	0.22408
MNN FOREST	-0.06993	0.33129	0.14555	0.63598	0.57415	0.07073	-0.14669	0.27653	-0.12896
IJI FOREST	0.53063	0.55632	-0.30954	0.16097	-0.14440	0.20024	-0.24833	0.06102	0.30708
LAND_NumP FOREST	0.74693	-0.11075	-0.38572	-0.24942	0.35865	-0.15188	-0.06989	-0.14364	0.05533
LAND_PSCoV FOREST	0.86333	-0.05233	-0.05137	0.01450	0.01828	-0.00291	-0.37726	-0.04137	-0.17762
LAND_PSSD FOREST	0.67437	-0.50830	0.34372	0.27892	-0.13741	0.07928	0.02133	-0.00453	0.08863
LAND_TE FOREST	0.83524	-0.40842	-0.09255	-0.07921	0.27195	-0.10883	0.07639	-0.08925	0.01772
LAND_ED FOREST	0.71851	0.25201	-0.42136	-0.19470	-0.20448	0.13512	-0.14362	0.08973	-0.05498
LAND_MSI FOREST	0.44042	0.52133	0.53686	-0.29219	-0.00856	-0.15924	0.06278	0.00843	-0.15528
LAND_AWMSI FOREST	0.82271	0.08773	0.41612	-0.13866	-0.05677	-0.14261	0.04577	-0.00772	-0.14110
LAND_MPFD FOREST	0.53250	0.76068	0.19441	0.09077	-0.03740	-0.20016	0.03762	-0.01143	0.12986
LAND_AWMPFD FOREST	0.56890	0.73069	0.18012	0.08013	-0.03819	-0.19481	0.07535	-0.00005	0.14693
LPI FOREST	0.78505	-0.22872	0.07975	0.11791	-0.34070	0.17476	-0.15596	0.19561	-0.24084
LSI FOREST	0.91859	-0.06512	-0.28727	-0.11715	0.16572	-0.02807	-0.01103	-0.01131	-0.00225
LAND_CA FOREST	0.79772	-0.51662	0.12012	0.09675	0.13983	-0.04905	0.08889	-0.07661	0.08019
DLFD FOREST	0.34752	0.47847	-0.14945	0.31094	0.00285	0.39532	0.28563	-0.50993	-0.18194
LAND_MPAR FOREST	0.05028	0.09682	0.50063	-0.53543	0.28471	0.58067	-0.06874	0.05840	0.12384

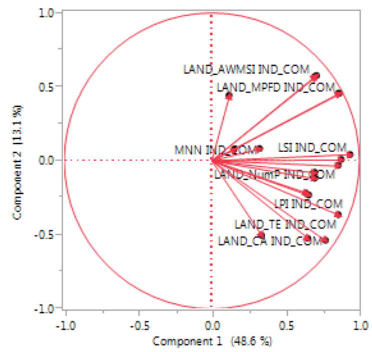
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS FOREST	-0.03771	-0.09989	0.04560	-0.01889	0.04519	-0.02927	0.01653	0.00037	0.00080
MPI FOREST	0.13824	0.09870	0.16153	-0.14761	0.00947	0.00158	0.00416	-0.00071	0.00245
MNN FOREST	0.07516	0.06404	-0.02344	-0.00107	0.00435	0.00873	0.00568	0.00012	-0.00032
IJI FOREST	0.07932	-0.21222	-0.10798	-0.04267	-0.02125	0.03585	0.00521	0.00189	0.00095
LAND_NumP FOREST	0.00971	0.00772	0.08304	0.02200	0.09316	0.13712	0.05917	0.00075	-0.00749
LAND_PSCoV FOREST	-0.12301	-0.14686	0.11843	-0.03806	0.06246	-0.12994	0.01114	-0.00680	0.00407
LAND_PSSD FOREST	0.01522	-0.03240	-0.05764	0.18416	0.13384	0.00773	-0.02810	0.00157	0.01162
LAND_TE FOREST	-0.00173	-0.01871	-0.05932	-0.00139	-0.11964	-0.01589	0.00468	0.00675	0.04789
LAND_ED FOREST	0.18824	0.25755	-0.03355	0.05573	0.00716	-0.07469	0.03704	0.00008	0.00302
LAND_MSI FOREST	0.27347	-0.15217	0.07344	0.04555	-0.03527	0.01750	-0.00912	-0.00729	-0.00048
LAND_AWMSI FOREST	-0.03233	0.04197	-0.21390	-0.14319	0.08778	0.01541	0.01794	0.01059	-0.00383
LAND_MPFD FOREST	-0.12213	0.06150	0.08287	0.05010	-0.02729	-0.02328	0.00160	0.05021	-0.00361
LAND_AWMPFD FOREST	-0.15145	0.10415	0.00691	0.01762	-0.00951	0.00721	0.00140	-0.05061	0.00758
LPI FOREST	-0.11431	0.02348	0.04022	0.01787	-0.10230	0.13361	0.00634	0.00076	-0.00207
LSI FOREST	0.01245	0.04949	-0.00510	-0.02051	-0.01156	0.00482	-0.15398	-0.00081	-0.01688
LAND_CA FOREST	-0.00680	-0.03936	-0.07060	0.05651	-0.10944	-0.06715	0.05332	-0.00690	-0.04063
DLFD FOREST	0.00301	-0.01217	0.00471	-0.00476	0.00388	0.00607	0.00364	-0.00015	0.00016
LAND_MPAR FOREST	-0.08435	0.04100	0.01304	0.00756	-0.00412	-0.00565	0.00067	0.00072	0.00046

Figure 53. Eigen values and Loading Matrix for Level 1 Forest Index

PCA Index – IND/COM

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	8.7430	48.572		48.572
2	2.3539	13.077		61.650
3	1.5616	8.676		70.325
4	1.2012	6.673		76.998
5	0.9744	5.413		82.412
6	0.7715	4.286		86.698
7	0.6947	3.860		90.557
8	0.5594	3.108		93.665
9	0.4183	2.324		95.989
10	0.2565	1.425		97.414
11	0.1815	1.008		98.423
12	0.1003	0.557		98.980
13	0.0684	0.380		99.360
14	0.0493	0.274		99.634
15	0.0297	0.165		99.799
16	0.0203	0.113		99.912
17	0.0154	0.086		99.997
18	0.0005	0.003		100.000



Loading Matrix

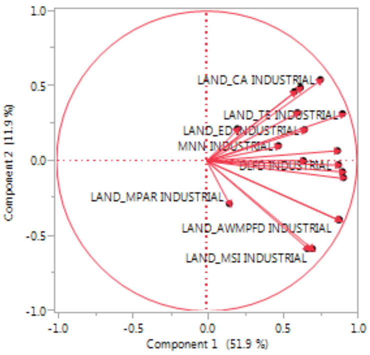
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS IND_COM	0.64232	-0.21825	-0.49638	0.06398	-0.01982	0.23643	0.28146	-0.31789	-0.08194	-0.13588	0.13205	0.04170	-0.06227	0.05118	0.01114	0.04061	0.01044	0.00019
MPI IND_COM	0.34226	-0.50948	0.28290	0.48536	0.16642	-0.12089	0.29258	0.26292	0.30181	-0.03691	0.13379	-0.00064	0.00783	-0.00391	0.00745	0.00281	0.01221	0.00002
MNN IND_COM	0.32431	0.07981	0.37708	-0.54964	-0.15459	0.44002	0.42764	0.17937	0.01447	0.10546	-0.00811	-0.01387	0.00125	-0.00345	-0.00105	-0.00048	-0.00018	-0.00012
IJI IND_COM	0.88285	0.00611	-0.17541	-0.23517	-0.03215	-0.03010	-0.04091	-0.05464	0.25244	-0.08187	-0.08411	0.16633	0.14688	0.03193	-0.00260	-0.00404	-0.00845	0.00001
LAND_NumP IND_COM	0.85886	-0.03547	0.28232	-0.18062	-0.01676	-0.11380	-0.23232	-0.13983	0.05942	0.14109	0.11478	-0.11756	0.06481	0.02190	0.00820	0.08266	0.00254	-0.00017
LAND_PSCoV IND_COM	0.69773	-0.12114	0.45635	-0.25776	-0.11527	-0.00473	-0.29011	0.10914	-0.12189	-0.25129	0.17647	0.03764	-0.04173	0.01092	0.02008	-0.03523	-0.01869	0.00056
LAND_PSSD IND_COM	0.65476	-0.52568	0.34733	0.17737	0.01350	0.04210	-0.04054	0.11210	-0.21114	-0.09430	-0.26168	0.02593	-0.01845	0.02692	0.01062	0.04624	0.02741	0.00016
LAND_TE IND_COM	0.85937	-0.36637	0.10476	-0.00422	0.01265	-0.01953	-0.06187	-0.23150	-0.05835	0.20724	0.03170	0.03092	-0.00196	-0.01228	0.01134	-0.07051	0.06877	-0.00001
LAND_ED IND_COM	0.70229	-0.08159	-0.47925	-0.05227	-0.06410	0.01348	-0.22991	0.38397	0.03334	0.20614	0.02910	0.07659	-0.09979	0.04846	-0.01543	0.01481	-0.00767	0.00002
LAND_MSI IND_COM	0.71271	0.57218	0.09605	0.17877	-0.00614	-0.13912	0.13431	0.08359	-0.23129	-0.01144	0.07672	0.06111	0.03976	-0.04321	-0.10819	0.01589	0.01975	0.00272
LAND_AWMSI IND_COM	0.69711	0.56719	0.03484	0.23715	0.02100	-0.17227	0.18739	0.05397	-0.20022	0.09858	0.01049	0.04891	0.02785	-0.00069	0.11350	-0.00206	-0.02732	-0.00253
LAND_MPFD IND_COM	0.86625	0.45080	0.00779	0.03086	0.00354	-0.06290	0.03337	-0.01668	0.11488	-0.08348	-0.06748	-0.09189	-0.04455	0.05515	-0.03090	-0.02305	0.01045	-0.01503
LAND_AWMPFD IND_COM	0.86008	0.45298	-0.01438	0.04129	0.01108	-0.07035	0.05249	-0.02695	0.13119	-0.05022	-0.08657	-0.10244	-0.04385	0.06328	0.00614	-0.02638	0.00111	0.01468
LPI IND_COM	0.66221	-0.23497	-0.58090	0.05666	-0.07642	0.19791	-0.05847	0.23983	-0.12339	-0.08974	0.01405	-0.14370	0.10882	-0.05395	0.01578	-0.02026	0.01076	-0.00008
LSI IND_COM	0.94171	0.04198	-0.08161	-0.13536	-0.00023	-0.07226	-0.01199	-0.09187	0.16561	-0.04137	-0.06717	0.01300	-0.09172	-0.17237	0.01416	0.02032	-0.00854	0.00003
LAND_CA IND_COM	0.77214	-0.54058	0.08560	0.19101	0.04590	0.04589	0.06455	-0.16436	-0.08500	0.10734	-0.03812	-0.02969	0.00829	0.00349	-0.05040	-0.02980	-0.08820	-0.00028
DLFD IND_COM	0.15365	0.07500	-0.03697	-0.23916	0.94584	0.10423	-0.03858	0.04077	-0.05798	-0.01169	0.00822	0.00149	0.00069	0.00151	0.00039	-0.00109	-0.00047	0.00001
LAND_MPAR IND_COM	0.12064	0.44140	0.22244	0.51454	0.01327	0.61305	-0.28751	-0.05917	0.10879	0.02703	0.01065	0.03422	0.00755	-0.01612	0.00461	0.00335	-0.00090	0.00007

Figure 54. Eigen values and Loading Matrix for Level 1 Ind/Com Index

PCA Index - INDUSTRIAL

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	9.3455	51.920	51.920
2	2.1454	11.919	63.838
3	1.6613	9.229	73.068
4	1.1501	6.389	79.457
5	0.9933	5.518	84.975
6	0.7242	4.023	88.998
7	0.6385	3.547	92.546
8	0.5247	2.915	95.461
9	0.2552	1.418	96.879
10	0.1736	0.965	97.843
11	0.1197	0.665	98.508
12	0.1099	0.610	99.119
13	0.0621	0.345	99.464
14	0.0431	0.239	99.703
15	0.0266	0.148	99.851
16	0.0173	0.096	99.947
17	0.0093	0.052	99.999
18	0.0003	0.001	100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS INDUSTRIAL	0.62580	0.48927	-0.37002	0.04946	0.07946	0.18012	-0.38452	0.12590	0.12712
MPI INDUSTRIAL	0.20793	0.21263	-0.03055	0.83589	0.28184	-0.06147	0.31177	0.14931	0.04757
MNN INDUSTRIAL	0.48190	0.09755	0.55984	-0.13939	-0.24424	-0.14302	-0.03236	0.58335	-0.04026
IJI INDUSTRIAL	0.87538	0.06950	-0.03554	-0.02732	-0.12116	-0.32985	-0.08005	-0.02921	0.01537
LAND_NumP INDUSTRIAL	0.91380	-0.11632	0.25773	0.01513	-0.04539	-0.07703	0.00316	-0.10441	0.10854
LAND_PSCoV INDUSTRIAL	0.64543	0.00249	0.61733	-0.14235	0.15226	0.04255	0.11017	-0.15455	0.32988
LAND_PSSD INDUSTRIAL	0.60945	0.32301	0.53947	-0.07918	0.05657	0.30346	0.14443	-0.15751	-0.25727
LAND_TE INDUSTRIAL	0.90369	0.31029	0.13145	0.06845	0.01443	0.08369	-0.01752	-0.08965	-0.06779
LAND_ED INDUSTRIAL	0.65357	0.20719	-0.37072	-0.36638	0.04771	-0.12901	0.45420	0.02670	-0.05487
LAND_MSI INDUSTRIAL	0.70693	-0.58966	-0.10952	-0.00437	0.03755	0.29153	0.09074	0.10632	0.08432
LAND_AWMSI INDUSTRIAL	0.67499	-0.58510	-0.20147	0.02011	0.00689	0.33535	0.09103	0.13510	-0.04785
LAND_MPFD INDUSTRIAL	0.88558	-0.39554	-0.11511	0.04464	-0.08261	-0.00744	-0.07515	-0.02825	-0.01581
LAND_AWMPFD INDUSTRIA	0.88035	-0.39524	-0.13758	0.04844	-0.07667	-0.00899	-0.08382	-0.01961	-0.05270
LPI INDUSTRIAL	0.58502	0.45932	-0.45717	-0.36356	0.08533	0.00052	0.17997	0.08142	0.09357
LSI INDUSTRIAL	0.90576	-0.07212	-0.06133	0.03844	-0.13202	-0.21584	-0.13467	-0.17081	-0.06575
LAND_CA INDUSTRIAL	0.75765	0.53722	-0.04367	0.10667	0.10198	0.25046	-0.18979	0.01527	-0.04475
DLFD INDUSTRIAL	0.87823	-0.02700	-0.07189	0.25858	-0.05547	-0.29260	0.02257	-0.05883	-0.06324
LAND_MPAR INDUSTRIAL	0.15070	-0.28483	0.10420	-0.21541	0.86599	-0.20664	-0.17359	0.08256	-0.09518

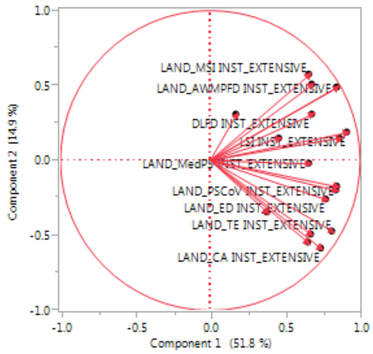
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS INDUSTRIAL	0.02200	-0.01937	0.03622	0.07791	0.01354	0.01233	0.00059	0.04921	0.00043
MPI INDUSTRIAL	0.02421	0.07088	0.00355	-0.00642	0.03559	-0.00044	0.00242	0.00378	0.00005
MNN INDUSTRIAL	0.03827	-0.01629	-0.01960	-0.00031	0.01451	-0.00108	-0.00319	-0.00242	0.00003
IJI INDUSTRIAL	-0.19975	0.14036	0.18838	-0.01668	-0.02188	-0.00001	0.00491	-0.00469	-0.00003
LAND_NumP INDUSTRIAL	-0.12389	0.00271	-0.17350	-0.05306	-0.00842	-0.00057	0.06970	0.02990	0.00036
LAND_PSCoV INDUSTRIAL	0.07204	-0.01447	0.04963	0.05172	0.00255	0.03041	-0.02135	-0.01695	-0.00081
LAND_PSSD INDUSTRIAL	0.08392	0.07351	0.07309	0.01703	-0.01986	-0.01127	0.02011	0.02804	0.00045
LAND_TE INDUSTRIAL	-0.14361	-0.06544	-0.07191	-0.07662	0.01748	0.00135	-0.09286	0.00793	0.00032
LAND_ED INDUSTRIAL	-0.08958	-0.05991	-0.03020	0.11979	0.04901	-0.00142	0.00480	0.00214	-0.00001
LAND_MSI INDUSTRIAL	-0.02843	-0.07555	0.08887	-0.03319	0.00718	-0.10902	0.00255	0.00530	-0.00092
LAND_AWMSI INDUSTRIAL	-0.04847	-0.02618	0.05000	-0.03452	-0.00490	0.11329	0.01077	-0.00365	0.00087
LAND_MPFD INDUSTRIAL	0.07574	0.11708	-0.07772	0.06080	-0.01178	-0.01911	-0.02276	-0.01241	0.01133
LAND_AWMPFD INDUSTRIA	0.06901	0.12425	-0.08581	0.05484	-0.01105	-0.00276	-0.02240	-0.00334	-0.01141
LPI INDUSTRIAL	0.15761	0.08111	-0.01687	-0.12980	-0.04448	-0.00115	-0.00034	-0.00354	-0.00002
LSI INDUSTRIAL	0.15292	-0.06858	0.04979	-0.05386	0.14006	0.00850	0.01774	-0.00580	0.00003
LAND_CA INDUSTRIAL	-0.06206	-0.03221	-0.03344	0.02050	-0.00785	-0.01489	0.03759	-0.06763	-0.00073
DLFD INDUSTRIAL	0.09569	-0.18161	0.03058	0.02087	-0.12545	0.00534	0.00158	0.00301	-0.00006
LAND_MPAR INDUSTRIAL	-0.00819	-0.00461	-0.00750	-0.00992	0.00187	-0.00062	0.00038	0.00132	0.00027

Figure 55. Eigen values and Loading Matrix for Level 1 Industrial Index

PCA Index – INST_EXTENSIVE

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	9.3250	51.806	51.806
2	2.6848	14.916	66.721
3	1.4277	7.932	74.653
4	1.0521	5.845	80.498
5	0.8285	4.603	85.101
6	0.6139	3.411	88.512
7	0.5462	3.035	91.546
8	0.4334	2.408	93.954
9	0.3827	2.126	96.080
10	0.2493	1.385	97.465
11	0.1639	0.910	98.376
12	0.1066	0.592	98.968
13	0.0828	0.460	99.428
14	0.0460	0.256	99.684
15	0.0254	0.141	99.825
16	0.0192	0.107	99.931
17	0.0119	0.066	99.997
18	0.0005	0.003	100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS INST_EXTENSIVE	0.66330	-0.02316	-0.26744	-0.38543	-0.08727	0.28804	0.34561	-0.25718	-0.19238
MPI INST_EXTENSIVE	0.38111	-0.34513	0.18184	-0.02293	0.79311	0.21780	0.10580	0.04803	0.04892
MNN INST_EXTENSIVE	0.45885	0.14812	-0.42546	0.57508	-0.07441	0.10740	0.40586	0.24902	0.06752
LII INST_EXTENSIVE	0.86628	0.14915	-0.33667	-0.02328	-0.05568	0.06836	-0.14272	-0.02116	-0.08188
LAND_NumP INST_EXTENSIVE	0.84977	-0.17593	-0.01836	0.26861	0.07341	-0.01883	-0.25731	0.02597	-0.21263
LAND_PSCoV INST_EXTENSIVE	0.84280	-0.20070	0.02001	0.30647	-0.06481	0.01546	-0.11833	0.12353	-0.08738
LAND_PSSD INST_EXTENSIVE	0.65614	-0.54994	0.19962	0.10203	-0.20368	-0.19309	0.15274	-0.17850	0.17754
LAND_TE INST_EXTENSIVE	0.81393	-0.47087	0.11462	0.16411	0.06744	-0.07453	-0.02933	-0.12944	-0.08032
LAND_ED INST_EXTENSIVE	0.77282	-0.26128	-0.00879	-0.36972	-0.10443	0.07643	-0.13854	0.34429	0.04631
LAND_MSI INST_EXTENSIVE	0.66255	0.57175	0.38174	-0.05531	0.04805	-0.19175	0.13308	0.07048	-0.00017
LAND_AWMSI INST_EXTENSIVE	0.68263	0.50540	0.37440	-0.09390	0.07755	-0.26502	0.17047	0.04725	0.03049
LAND_MPFD INST_EXTENSIVE	0.84717	0.48495	0.01387	-0.05138	0.01884	-0.03132	-0.03066	-0.01687	-0.05459
LAND_AWMPFD INST_EXTENSIVE	0.84509	0.48685	0.00952	-0.06475	0.02660	-0.03528	-0.02432	-0.02437	-0.05204
LPI INST_EXTENSIVE	0.67187	-0.49038	0.08609	-0.37645	-0.18979	0.08655	0.07526	0.22168	0.13770
LSI INST_EXTENSIVE	0.91378	0.18486	-0.22807	-0.07382	0.02154	0.00866	-0.09679	-0.07322	-0.04657
LAND_CA INST_EXTENSIVE	0.73677	-0.58474	0.17189	0.11244	-0.06905	-0.11871	0.05055	-0.15025	0.02254
DLFD INST_EXTENSIVE	0.67785	0.30402	-0.36538	0.02319	0.09848	0.06078	-0.18730	-0.17321	0.45492
LAND_MPAR INST_EXTENSIVE	0.17504	0.30633	0.68285	0.22936	-0.23739	0.53243	-0.07693	-0.06526	0.06239

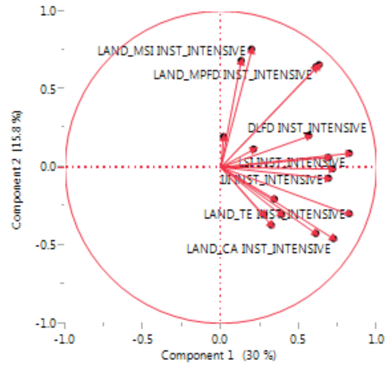
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS INST_EXTENSIVE	0.11586	-0.10970	0.00576	-0.01207	0.01871	0.00896	0.01304	0.00059	0.00007
MPI INST_EXTENSIVE	-0.09625	0.00194	0.00489	0.01515	0.00486	0.00455	0.01092	0.00118	0.00003
MNN INST_EXTENSIVE	0.03325	0.07613	0.01184	-0.01817	-0.00618	0.00127	0.00088	-0.00058	0.00008
LII INST_EXTENSIVE	-0.13511	0.08661	0.07669	0.20928	0.05084	-0.00733	-0.00648	0.00281	0.00019
LAND_NumP INST_EXTENSIVE	0.20294	0.03185	-0.07480	0.01342	-0.01769	0.07599	0.05062	-0.00187	0.00031
LAND_PSCoV INST_EXTENSIVE	-0.10522	-0.31649	-0.02912	-0.00132	0.01502	-0.03307	-0.01618	-0.00892	-0.00072
LAND_PSSD INST_EXTENSIVE	-0.20049	-0.02130	0.05987	-0.01431	0.01612	0.07356	0.02844	0.01971	-0.00019
LAND_TE INST_EXTENSIVE	0.14330	0.07050	0.06357	-0.03642	0.00792	-0.00460	-0.10005	0.02440	-0.00012
LAND_ED INST_EXTENSIVE	0.06771	0.02110	0.15680	-0.09088	0.04974	-0.00794	0.02557	0.00466	0.00003
LAND_MSI INST_EXTENSIVE	0.09633	-0.04880	0.00735	0.06229	-0.03886	-0.03301	0.02531	0.06896	-0.00230
LAND_AWMSI INST_EXTENSIVE	0.08921	-0.03548	0.06150	0.04558	0.00748	0.02996	-0.01895	-0.06888	0.00229
LAND_MPFD INST_EXTENSIVE	-0.12609	0.05871	-0.09661	-0.08509	0.04848	-0.00520	-0.00548	0.01349	0.01595
LAND_AWMPFD INST_EXTENSIVE	-0.11707	0.07831	-0.09025	-0.08465	0.05596	0.00440	-0.00900	-0.00679	-0.01586
LPI INST_EXTENSIVE	-0.00365	0.03145	-0.17757	0.07001	-0.05368	0.01666	-0.03341	-0.00200	0.00000
LSI INST_EXTENSIVE	-0.12912	0.02892	0.07580	-0.05506	-0.17127	-0.01143	0.00152	-0.01332	-0.00008
LAND_CA INST_EXTENSIVE	0.05303	0.10102	-0.04247	0.00488	0.01219	-0.10198	0.04859	-0.02840	0.00025
DLFD INST_EXTENSIVE	0.15270	-0.06825	-0.02140	-0.00274	0.01325	-0.00143	-0.00019	0.00123	0.00001
LAND_MPAR INST_EXTENSIVE	-0.00380	0.03965	0.03116	0.00582	-0.00963	0.00271	-0.00094	-0.00600	0.00011

Figure 56. Eigen values and Loading Matrix for Level 1 Institutional Extensive Index

PCA Index INST_INTENSIVE

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	5.3918	29.954	29.954
2	2.8364	15.758	45.712
3	2.1548	11.971	57.683
4	1.6591	9.217	66.900
5	1.0536	5.853	72.754
6	0.9913	5.507	78.261
7	0.7884	4.380	82.641
8	0.7595	4.220	86.861
9	0.6065	3.369	90.230
10	0.5115	2.842	93.072
11	0.3947	2.193	95.265
12	0.3465	1.925	97.190
13	0.2349	1.305	98.495
14	0.1311	0.728	99.223
15	0.0727	0.404	99.627
16	0.0445	0.247	99.874
17	0.0169	0.094	99.968
18	0.0058	0.032	100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS INST_INTENSIVE	0.27723	-0.30941	0.48234	-0.41972	0.45105	0.04753	-0.27279	0.02828	0.01519
MPI INST_INTENSIVE	0.39053	-0.30806	0.24430	0.32089	-0.17839	-0.07671	0.29168	-0.58663	0.18383
MNN INST_INTENSIVE	0.21078	0.11352	-0.40810	-0.23989	0.43914	0.24795	0.59827	0.18679	0.22525
IJI INST_INTENSIVE	0.69294	-0.07274	-0.20867	-0.34958	-0.01868	0.07649	-0.22793	0.07806	0.17928
LAND_NumP INST_INTENSIVE	0.71829	-0.01138	-0.41276	0.17856	-0.28625	-0.04282	-0.14884	0.19553	0.20431
LAND_PSCoV INST_INTENSIVE	0.69413	0.06271	-0.30758	0.21988	-0.26900	0.08582	0.19287	0.23180	-0.26475
LAND_PSSD INST_INTENSIVE	0.61356	-0.42737	0.14496	0.39600	0.32276	0.02266	0.11239	-0.01593	-0.27281
LAND_TE INST_INTENSIVE	0.82398	-0.30242	-0.20064	0.24858	0.02838	-0.04605	-0.19105	0.08018	0.18095
LAND_ED INST_INTENSIVE	0.34335	-0.20406	0.63734	-0.23550	-0.46451	0.04461	0.17013	0.18793	0.12939
LAND_MSI INST_INTENSIVE	0.19992	0.75484	0.35456	0.39130	0.04697	0.00794	0.02171	0.11838	0.13525
LAND_AWMSI INST_INTENSIVE	0.13043	0.68158	0.43149	0.38060	0.21895	-0.06368	-0.07458	0.06179	0.30011
LAND_MPFD INST_INTENSIVE	0.63185	0.65562	0.10070	-0.14093	-0.02538	-0.10477	0.05448	-0.03522	-0.29282
LAND_AWMPFD INST_INTENSIVE	0.61181	0.63950	0.13723	-0.18930	0.06655	-0.11999	-0.01648	-0.09320	-0.20367
LPI INST_INTENSIVE	0.32797	-0.37515	0.73719	-0.18919	-0.09560	0.09914	0.16928	0.23286	-0.03934
LSI INST_INTENSIVE	0.82760	0.08862	-0.12587	-0.28885	-0.03903	-0.01444	-0.07445	-0.15474	-0.00208
LAND_CA INST_INTENSIVE	0.72444	-0.45871	0.03229	0.37215	0.27550	-0.01646	-0.06836	-0.00773	-0.05075
DLFD INST_INTENSIVE	0.56450	0.19874	-0.12016	-0.45731	-0.00289	0.02411	0.06408	-0.33116	0.10442
LAND_MPAR INST_INTENSIVE	0.01670	0.19009	0.06564	0.15936	-0.09116	0.92825	-0.17145	-0.15494	-0.05659

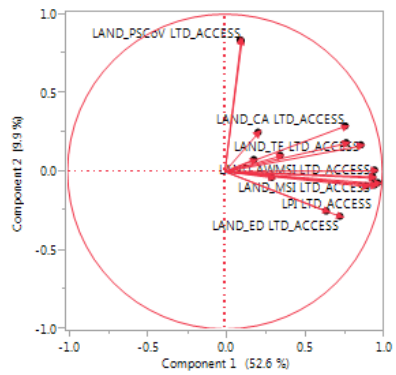
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS INST_INTENSIVE	0.16595	-0.00173	0.31163	0.01453	0.10658	0.01819	0.02962	0.00310	0.00093
MPI INST_INTENSIVE	0.25700	0.11156	0.08958	0.05182	0.00212	0.00009	0.00399	-0.00027	-0.00004
MNN INST_INTENSIVE	0.12696	-0.05475	-0.03204	-0.01184	0.01876	-0.00862	0.00257	0.00073	-0.00066
IJI INST_INTENSIVE	0.04670	0.44613	-0.19131	0.11786	-0.01838	-0.01812	0.00516	0.00341	-0.00150
LAND_NumP INST_INTENSIVE	0.11087	-0.21158	0.11055	0.08245	-0.07676	-0.00149	0.11506	0.02366	0.00309
LAND_PSCoV INST_INTENSIVE	-0.04957	0.19571	0.27435	-0.00002	0.05984	0.09215	-0.04861	0.00638	0.00675
LAND_PSSD INST_INTENSIVE	-0.18936	0.07127	-0.12342	-0.01146	-0.00160	0.01946	0.11230	-0.04206	-0.00334
LAND_TE INST_INTENSIVE	0.03538	-0.17037	-0.02284	0.01734	0.04929	-0.02158	-0.09427	-0.07730	-0.00976
LAND_ED INST_INTENSIVE	-0.06540	-0.07828	-0.16819	-0.05804	0.19227	0.01720	0.02568	0.00710	-0.00087
LAND_MSI INST_INTENSIVE	-0.11139	0.12430	0.12841	-0.05080	0.03387	-0.16629	0.01196	-0.01013	0.02002
LAND_AWMSI INST_INTENSIVE	-0.07130	0.03466	-0.03698	-0.02060	-0.04252	0.15983	-0.00876	0.01167	-0.01791
LAND_MPFD INST_INTENSIVE	0.15164	-0.05666	-0.02500	0.07331	0.01835	-0.06367	-0.00005	0.01105	-0.05136
LAND_AWMPFD INST_INTENSIVE	0.19633	-0.12172	-0.15733	0.09702	-0.00042	0.05253	-0.01000	-0.01422	0.04692
LPI INST_INTENSIVE	-0.00693	-0.02390	0.05214	0.06052	-0.24490	-0.02090	-0.03984	-0.00756	0.00108
LSI INST_INTENSIVE	0.02102	0.00975	-0.00210	-0.41248	-0.07757	0.00237	0.00306	0.00336	0.00081
LAND_CA INST_INTENSIVE	-0.07720	-0.08783	-0.11638	0.01113	0.02910	-0.04272	-0.06216	0.08841	0.00733
DLFD INST_INTENSIVE	-0.50587	-0.09861	0.10820	0.12491	-0.00490	0.00457	-0.00422	0.00044	0.00036
LAND_MPAR INST_INTENSIVE	0.04559	-0.06422	-0.03180	0.00863	-0.00021	0.00529	0.00016	-0.00059	-0.00171

Figure 57. Eigen values and Loading Matrix for Level 1 Institutional Intensive Index

PCA Index – LTD_ACCESS

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	9.4626	52.570		52.570
2	1.7825	9.903		62.473
3	1.4507	8.059		70.532
4	1.1863	6.591		77.123
5	1.1362	6.312		83.436
6	0.8915	4.953		88.388
7	0.7523	4.179		92.567
8	0.5314	2.952		95.520
9	0.3305	1.836		97.356
10	0.2314	1.286		98.642
11	0.1411	0.784		99.426
12	0.0708	0.393		99.819
13	0.0188	0.104		99.923
14	0.0132	0.073		99.997
15	0.0006	0.003		100.000
16	0.0000	0.000		100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS LTD_ACCESS	0.77157	0.17842	-0.53477	0.00125	-0.01254	-0.17084	-0.00953	0.21653	-0.04850
MPI LTD_ACCESS	0.30248	-0.03916	-0.20584	-0.01086	-0.38066	0.83989	0.05943	0.06307	-0.02224
MNN LTD_ACCESS	0.21410	0.24704	0.19632	0.62753	0.29838	0.16105	-0.55710	0.15547	0.09913
IJI LTD_ACCESS	0.89848	-0.09462	0.08081	-0.04356	0.14588	0.05710	-0.06367	-0.13168	-0.13343
LAND_NumP LTD_ACCESS	0.95014	0.00498	0.21027	-0.02483	-0.01850	-0.04155	-0.03823	-0.14129	-0.08463
LAND_PSCoV LTD_ACCESS	0.10850	0.82911	0.32423	-0.15581	0.08666	0.10535	0.02835	0.08461	-0.34668
LAND_PSSD LTD_ACCESS	0.10270	0.83454	0.04776	-0.26559	0.15100	0.08803	0.12009	-0.22497	0.33927
LAND_TE LTD_ACCESS	0.86936	0.16336	-0.39152	0.06281	-0.12581	-0.05422	-0.02019	0.04709	0.04028
LAND_ED LTD_ACCESS	0.73146	-0.28533	0.20709	-0.19525	0.40519	0.18187	0.09603	0.11295	0.13457
LAND_MSI LTD_ACCESS	0.94188	-0.04169	0.03021	0.09331	-0.21769	-0.02368	-0.01644	-0.08260	0.09179
LAND_AWMSI LTD_ACCESS	0.94267	-0.03350	0.03266	0.09339	-0.21525	-0.02195	-0.01875	-0.08264	0.09117
LAND_MPFD LTD_ACCESS	0.95452	-0.08587	0.19198	-0.00443	-0.06756	-0.05853	-0.02245	-0.13723	-0.03417
LAND_AWMPFD LTD_ACCESS	0.95447	-0.08890	0.18950	-0.00306	-0.06668	-0.05852	-0.02405	-0.13916	-0.02885
LPI LTD_ACCESS	0.64902	-0.25447	0.04371	-0.31047	0.53010	0.10588	0.12443	0.27297	0.01863
LSI LTD_ACCESS	0.97164	-0.07588	0.03956	-0.02992	0.04137	-0.02114	-0.05072	-0.10734	-0.06686
LAND_CA LTD_ACCESS	0.76516	0.28849	-0.51435	-0.03454	0.00932	-0.15384	0.00599	0.17926	0.00106
DLFD LTD_ACCESS	0.18566	0.07649	-0.03095	0.73151	0.24884	0.03606	0.59483	-0.06837	-0.03957
LAND_MPAR LTD_ACCESS	0.35167	0.09723	0.62106	0.04411	-0.47196	-0.16306	0.18518	0.41425	0.11293

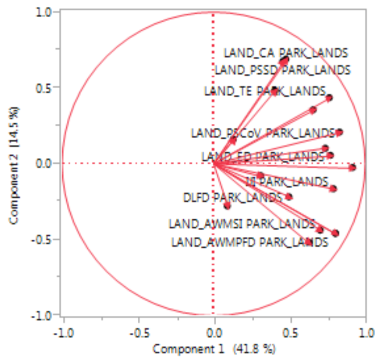
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS LTD_ACCESS	0.06927	0.00067	-0.05173	-0.00815	0.02567	0.00131	0.00059	0.00000	0.00000
MPI LTD_ACCESS	0.07769	-0.00954	-0.00522	-0.00226	0.00199	0.00003	-0.00000	-0.00000	0.00000
MNN LTD_ACCESS	0.03719	-0.00705	0.00617	-0.00123	-0.00012	0.00037	0.00000	0.00000	0.00000
IJI LTD_ACCESS	0.02715	0.33296	-0.00256	-0.00247	-0.00033	-0.00002	-0.00000	-0.00000	0.00000
LAND_NumP LTD_ACCESS	0.11432	-0.08045	-0.00746	-0.04018	-0.01275	-0.01871	0.00040	0.00000	0.00000
LAND_PSCoV LTD_ACCESS	-0.15011	-0.04514	0.00396	0.00024	0.00704	0.00202	-0.00005	-0.00000	0.00000
LAND_PSSD LTD_ACCESS	0.09109	0.03336	0.00760	0.00485	-0.00064	0.00025	0.00009	0.00000	0.00000
LAND_TE LTD_ACCESS	-0.16330	0.00589	0.04870	-0.00695	-0.09029	0.00078	-0.00002	-0.00000	0.00000
LAND_ED LTD_ACCESS	-0.19156	-0.02618	-0.16627	-0.01202	-0.00490	-0.00029	0.00001	0.00000	0.00000
LAND_MSI LTD_ACCESS	-0.18138	-0.00840	0.05853	0.00577	0.04002	0.00169	0.00124	-0.00000	0.00000
LAND_AWMSI LTD_ACCESS	-0.18155	-0.00899	0.05883	0.00477	0.04011	-0.00413	-0.00119	0.00000	0.00000
LAND_MPFD LTD_ACCESS	0.10131	-0.06755	-0.00860	-0.03281	-0.00453	0.01242	0.00062	0.00000	0.00000
LAND_AWMPFD LTD_ACCESS	0.10348	-0.06726	-0.00841	-0.03349	-0.00456	0.00842	-0.00105	-0.00000	0.00000
LPI LTD_ACCESS	0.06897	-0.03392	0.16502	0.00061	0.00434	0.00033	-0.00001	-0.00000	0.00000
LSI LTD_ACCESS	0.08325	-0.06648	-0.03291	0.12000	-0.01151	-0.00038	-0.00001	-0.00000	0.00000
LAND_CA LTD_ACCESS	0.08062	0.00519	-0.04916	-0.00776	0.02492	-0.00156	-0.00061	-0.00000	0.00000
DLFD LTD_ACCESS	0.01784	-0.00156	0.00143	0.00115	0.00035	-0.00006	-0.00000	-0.00000	0.00000
LAND_MPAR LTD_ACCESS	0.07667	0.07013	-0.01275	0.01145	-0.01193	-0.00069	0.00001	0.00000	0.00000

Figure 58. Eigen values and Loading Matrix for Level 1 Limited Access Index

PCA Index – PARK_LANDS

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	7.5157	41.754		41.754
2	2.6020	14.456		56.210
3	1.6205	9.003		65.212
4	1.2228	6.793		72.006
5	1.0799	6.000		78.005
6	0.8772	4.873		82.878
7	0.7671	4.262		87.140
8	0.6724	3.736		90.876
9	0.5811	3.229		94.104
10	0.3978	2.210		96.314
11	0.2438	1.354		97.669
12	0.1494	0.830		98.498
13	0.1232	0.685		99.183
14	0.0745	0.414		99.597
15	0.0356	0.198		99.794
16	0.0212	0.118		99.912
17	0.0153	0.085		99.997
18	0.0005	0.003		100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS PARK_LANDS	0.13436	0.16274	0.18682	0.56319	-0.68080	0.28588	-0.05620	0.00750	0.20547
MPI PARK_LANDS	0.40553	0.48123	0.20973	-0.15426	0.08762	-0.40442	0.30191	0.01730	0.47760
MNN PARK_LANDS	0.31144	-0.08316	-0.19942	0.44058	0.55833	0.42373	0.16434	-0.22102	0.27434
IJI PARK_LANDS	0.79312	-0.16565	-0.31589	0.22422	0.00833	-0.06518	0.14527	-0.02660	-0.20059
LAND_NumP PARK_LANDS	0.73819	0.10258	-0.32328	-0.37048	-0.07130	0.30383	-0.20420	0.04576	0.07345
LAND_PSCoV PARK_LANDS	0.83051	0.20552	-0.14281	-0.04419	0.17211	0.11762	0.09715	-0.11246	-0.11082
LAND_PSSD PARK_LANDS	0.45951	0.67030	0.38381	0.07046	0.32411	-0.06215	-0.15455	0.03739	-0.19817
LAND_TE PARK_LANDS	0.76800	0.43366	-0.07856	-0.30804	-0.07834	0.19102	-0.16897	0.06312	0.12873
LAND_ED PARK_LANDS	0.77027	0.05023	-0.33590	-0.23723	-0.24923	-0.00597	0.21791	-0.02490	-0.04599
LAND_MSI PARK_LANDS	0.63089	-0.52329	0.46793	-0.08289	-0.00186	-0.01473	-0.10096	-0.11961	0.08394
LAND_AWMSI PARK_LANDS	0.70686	-0.43820	0.42188	-0.16776	-0.03105	-0.03907	-0.12231	-0.09307	0.16276
LAND_MPFD PARK_LANDS	0.80898	-0.46209	0.11244	0.15157	0.02835	-0.14870	-0.09564	-0.07269	-0.08019
LAND_AWMPFD PARK_LANDS	0.80699	-0.46330	0.11282	0.14416	0.02208	-0.15144	-0.10503	-0.06536	-0.06890
LPI PARK_LANDS	0.66068	0.35362	0.06921	0.22622	-0.22232	-0.17546	0.40284	-0.11215	-0.17338
LSI PARK_LANDS	0.91873	-0.02990	-0.26104	-0.09114	-0.09926	-0.00044	-0.03017	0.02643	0.00160
LAND_CA PARK_LANDS	0.48302	0.68651	0.38338	0.20365	0.06367	0.09039	-0.23621	0.06438	-0.12306
DLFD PARK_LANDS	0.50191	-0.21809	-0.23391	0.31921	0.14928	-0.15876	-0.06665	0.66913	0.11041
LAND_MPAR PARK_LANDS	0.09521	-0.27950	0.56025	-0.24595	0.03270	0.43661	0.45265	0.32261	-0.13261

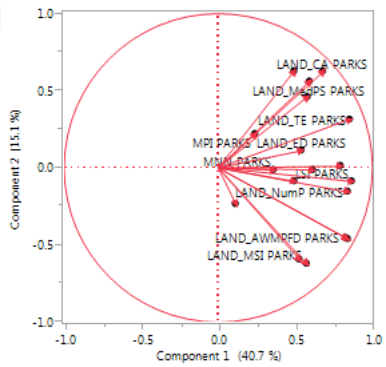
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS PARK_LANDS	0.07772	-0.05684	0.02433	-0.04400	0.00476	0.00276	-0.00300	0.02988	0.00005
MPI PARK_LANDS	0.20852	-0.01481	0.01304	-0.00240	0.03418	0.01891	-0.00019	-0.00064	0.00011
MNN PARK_LANDS	-0.08732	0.10908	-0.01546	0.00554	0.00123	-0.00210	0.00299	-0.00071	-0.00011
IJI PARK_LANDS	0.22315	-0.02304	0.22402	0.15904	0.02467	-0.02160	-0.01426	0.00364	-0.00011
LAND_NumP PARK_LANDS	0.08354	-0.05261	-0.12171	0.13557	0.09363	0.04811	0.04218	0.01992	-0.00025
LAND_PSCoV PARK_LANDS	0.01716	-0.38254	-0.00973	-0.14437	-0.00684	-0.01043	-0.00764	-0.00800	0.00070
LAND_PSSD PARK_LANDS	-0.03646	0.05872	0.03589	-0.01830	-0.01041	0.01086	0.00442	0.08465	-0.00006
LAND_TE PARK_LANDS	-0.01412	0.06333	-0.02455	0.05217	-0.09793	-0.07281	-0.08246	-0.01016	0.00023
LAND_ED PARK_LANDS	-0.22853	0.17036	0.07321	-0.14033	0.12642	-0.00725	-0.01497	0.01083	-0.00002
LAND_MSI PARK_LANDS	-0.20686	-0.06873	0.07541	0.05770	0.00106	0.10542	-0.05366	-0.00504	0.00245
LAND_AWMSI PARK_LANDS	-0.15932	-0.04786	0.07859	0.02872	-0.00590	-0.10020	0.06948	0.00348	-0.00260
LAND_MPFD PARK_LANDS	0.16728	0.06397	-0.12771	-0.05314	0.02143	0.00123	-0.02690	0.00031	-0.01554
LAND_AWMPFD PARK_LANDS	0.17266	0.08447	-0.12697	-0.04797	0.02425	-0.03019	-0.00446	0.00196	0.01549
LPI PARK_LANDS	-0.22907	-0.01138	-0.15899	0.12980	-0.04804	0.00028	0.01568	-0.00892	-0.00008
LSI PARK_LANDS	0.07163	0.11709	0.06367	-0.08690	-0.16997	0.06836	0.05472	-0.01169	0.00002
LAND_CA PARK_LANDS	0.01546	0.06818	0.04476	-0.01162	0.07615	0.01100	0.01915	-0.07953	-0.00015
DLFD PARK_LANDS	-0.16628	-0.06344	-0.01813	0.00234	0.00618	-0.00217	-0.00203	0.00003	-0.00000
LAND_MPAR PARK_LANDS	0.13539	0.03968	-0.00889	-0.01273	-0.00697	-0.00119	0.00132	-0.00018	-0.00004

Figure 59. Eigen values and Loading Matrix for Level 1 Park Lands Index

PCA Index – PARKS

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	7.3223	40.679	40.679
2	2.7257	15.143	55.822
3	1.6378	9.099	64.921
4	1.4360	7.978	72.899
5	1.1681	6.489	79.388
6	0.8647	4.804	84.192
7	0.7560	4.200	88.392
8	0.5980	3.322	91.714
9	0.5620	3.122	94.836
10	0.2846	1.581	96.418
11	0.1947	1.082	97.499
12	0.1709	0.950	98.449
13	0.1112	0.618	99.066
14	0.0833	0.463	99.529
15	0.0436	0.242	99.771
16	0.0314	0.174	99.946
17	0.0090	0.050	99.996
18	0.0008	0.004	100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MadPS PARKS	0.59245	0.56233	0.19003	0.36875	-0.26014	0.09613	0.03646	0.04109	-0.12994
MPI PARKS	0.24182	0.22318	0.29456	-0.50515	0.32354	-0.35207	0.43819	-0.22340	-0.26526
MNN PARKS	0.35960	-0.01356	-0.62313	0.04677	0.13929	-0.21901	0.25909	0.57478	-0.11974
LJI PARKS	0.79650	0.01087	-0.26917	-0.18477	-0.12288	0.18651	-0.15543	-0.02406	-0.23931
LAND_NumP PARKS	0.83760	-0.15569	-0.33267	-0.10424	0.18022	-0.05247	-0.13045	-0.16246	0.06347
LAND_PSCoV PARKS	0.61648	-0.01340	-0.35717	-0.09052	0.52595	-0.09171	-0.27399	-0.10592	0.21710
LAND_PSSD PARKS	0.49076	0.62555	0.11577	0.40597	0.19342	-0.13510	-0.02519	0.01527	0.17802
LAND_TE PARKS	0.85253	0.31554	-0.13495	0.05634	0.04708	-0.06845	0.06208	-0.16168	-0.02788
LAND_ED PARKS	0.54292	0.11198	0.35584	-0.60794	0.04535	0.13047	-0.13445	0.24816	0.21254
LAND_MSI PARKS	0.57326	-0.62322	0.34643	0.24406	0.18143	-0.10013	0.03309	0.05207	0.12957
LAND_AWMSI PARKS	0.52860	-0.59414	0.40393	0.24940	-0.05843	-0.28267	0.09646	0.06761	0.07603
LAND_MPFD PARKS	0.83911	-0.46089	0.09249	0.04775	-0.14864	-0.02247	-0.06290	-0.01151	-0.08790
LAND_AWMPFD PARKS	0.82910	-0.45461	0.10855	0.05068	-0.19786	-0.03580	-0.03661	-0.00210	-0.11550
LPI PARKS	0.57830	0.46016	0.45492	-0.32566	-0.03157	0.12217	-0.05812	0.23975	0.05944
LSI PARKS	0.86558	-0.08762	-0.15682	-0.13147	-0.22196	0.14873	-0.06225	-0.07155	-0.18717
LAND_CA PARKS	0.68132	0.62441	0.05656	0.35078	-0.00026	-0.04636	0.03577	-0.04915	0.01169
DLFD PARKS	0.49482	-0.08787	-0.29712	-0.12014	-0.31006	0.26046	0.53967	-0.12651	0.40767
LAND_MPAR PARKS	0.11303	-0.23230	0.17659	0.25395	0.60536	0.63154	0.20304	0.03827	-0.16079

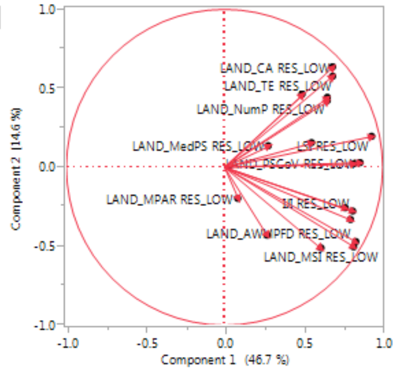
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MadPS PARKS	-0.02510	-0.09807	0.19112	0.01862	0.10721	0.04026	0.05423	0.02660	0.00017
MPI PARKS	0.08435	0.00570	0.01848	0.01116	0.03044	0.01621	0.00733	-0.00170	-0.00001
MNN PARKS	-0.01015	-0.02544	-0.00516	0.00242	0.00293	0.00452	-0.00313	-0.00041	0.00017
LJI PARKS	0.04961	0.34078	0.09029	-0.00832	0.00951	-0.00172	-0.00913	-0.00022	-0.00001
LAND_NumP PARKS	-0.15797	-0.02034	-0.11204	-0.12570	-0.00029	0.11304	0.05839	0.00840	0.00037
LAND_PSCoV PARKS	0.14819	-0.08423	0.17712	0.01488	0.01983	-0.06342	0.02114	-0.00592	-0.00114
LAND_PSSD PARKS	0.22376	0.11367	-0.18367	0.05797	-0.00862	0.01433	-0.00628	0.03186	0.00045
LAND_TE PARKS	-0.31491	-0.03013	-0.00556	0.02190	-0.04470	-0.07810	-0.06946	0.03249	0.00071
LAND_ED PARKS	-0.13691	0.02026	-0.07560	0.12769	0.10569	-0.00275	0.01436	-0.00509	0.00015
LAND_MSI PARKS	-0.02291	0.03025	0.13203	0.05691	-0.01808	0.10108	-0.08472	-0.00378	-0.00353
LAND_AWMSI PARKS	-0.08679	0.10864	0.02520	0.01208	-0.07586	-0.06343	0.09667	0.00028	0.00313
LAND_MPFD PARKS	0.11953	-0.08130	-0.06561	-0.05748	0.08424	-0.02431	-0.04032	-0.00362	0.01969
LAND_AWMPFD PARKS	0.09362	-0.06580	-0.10220	-0.06525	0.07532	-0.04962	-0.01139	0.00315	-0.01947
LPI PARKS	0.07877	-0.04904	0.05874	-0.17476	-0.13270	-0.00251	-0.01589	0.00505	-0.00009
LSI PARKS	0.10052	-0.14008	-0.04302	0.17996	-0.14332	0.02804	0.02113	-0.00783	-0.00003
LAND_CA PARKS	-0.07718	0.01314	-0.03388	-0.01633	0.01090	-0.00546	-0.00648	-0.07716	-0.00087
DLFD PARKS	0.07370	0.02401	0.03094	-0.01549	0.00832	-0.00452	0.00221	-0.00048	-0.00010
LAND_MPAR PARKS	-0.02481	-0.00808	-0.04751	-0.00791	0.00043	-0.02037	0.01799	0.00113	0.00047

Figure 60. Eigen values and Loading Matrix for Level 1 Parks Index

PCA Index – RES_LOW

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	8.4094	46.719		46.719
2	2.6221	14.567		61.286
3	1.8595	10.331		71.617
4	1.1448	6.360		77.977
5	0.9425	5.236		83.213
6	0.8232	4.573		87.786
7	0.7075	3.931		91.717
8	0.4371	2.428		94.145
9	0.2653	1.474		95.619
10	0.1983	1.101		96.720
11	0.1682	0.934		97.655
12	0.1493	0.829		98.484
13	0.1022	0.568		99.052
14	0.0822	0.457		99.508
15	0.0503	0.280		99.788
16	0.0311	0.173		99.961
17	0.0045	0.025		99.986
18	0.0026	0.014		100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS RES_LOW	0.27776	0.13126	0.65083	-0.04741	0.04203	0.66781	0.06356	0.07527	0.04305
MPI RES_LOW	0.49433	0.45979	0.34892	0.15161	0.01833	-0.31416	0.28690	0.44536	0.09122
MNN RES_LOW	0.27178	-0.43505	-0.14855	-0.19116	0.66238	0.01962	0.45661	-0.08596	0.12759
IJI RES_LOW	0.81046	-0.28123	0.07270	-0.26037	0.14643	-0.04030	-0.20735	0.08062	-0.03810
LAND_NumP RES_LOW	0.65560	0.41968	-0.53495	-0.01682	-0.00557	0.24264	-0.04753	-0.06254	0.04726
LAND_PSCoV RES_LOW	0.85758	0.02907	-0.10929	-0.01229	0.14028	-0.23365	-0.13872	-0.18579	0.11503
LAND_PSSD RES_LOW	0.65357	0.43708	0.35861	0.11675	0.10149	-0.30008	0.10875	-0.22348	-0.23040
LAND_TE RES_LOW	0.68467	0.57622	-0.39573	0.05429	-0.00636	0.16719	0.05934	-0.01516	-0.00019
LAND_ED RES_LOW	0.82089	0.01853	0.19333	-0.14233	-0.03916	-0.12885	-0.34377	0.03199	0.28054
LAND_MSI RES_LOW	0.61055	-0.51414	-0.00307	0.43933	-0.21619	0.04879	0.11772	-0.10279	0.10548
LAND_AWMSI RES_LOW	0.76440	-0.26187	0.04110	0.21906	-0.39339	-0.02906	0.24577	-0.08200	0.06841
LAND_MPFD RES_LOW	0.82187	-0.50547	-0.05964	-0.02636	-0.13800	0.03882	0.07070	0.04216	-0.09660
LAND_AWMPFD RES_LOW	0.83337	-0.47376	-0.05843	-0.03929	-0.15825	0.05073	0.07774	0.05763	-0.11189
LPI RES_LOW	0.55151	0.15178	0.74629	-0.01260	0.09971	0.03834	-0.07964	-0.23061	-0.02400
LSI RES_LOW	0.93363	0.19456	-0.16822	-0.10647	0.02423	0.02746	-0.10613	0.06444	0.01902
LAND_CA RES_LOW	0.68454	0.63159	-0.25247	0.10107	0.02207	0.10904	0.12919	-0.01314	-0.06455
DLFD RES_LOW	0.80180	-0.33537	-0.07232	-0.23620	0.10206	0.01931	-0.12227	0.19556	-0.21236
LAND_MPAR RES_LOW	0.08808	-0.19890	-0.05713	0.81071	0.42621	0.08174	-0.28299	0.10695	-0.04743

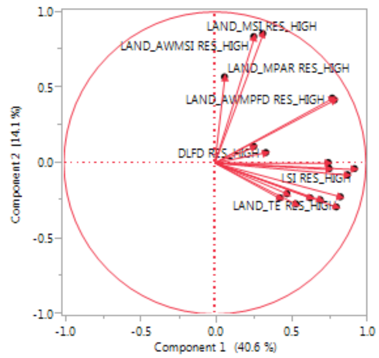
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS RES_LOW	-0.07791	0.02217	-0.03869	-0.09732	-0.04205	0.03087	0.01595	0.00115	-0.00037
MPI RES_LOW	-0.08298	0.00871	-0.02086	0.05542	0.02702	0.01668	-0.00287	0.00078	-0.00040
MNN RES_LOW	0.08089	0.00226	0.01945	0.00266	-0.00066	0.00572	-0.00039	-0.00032	-0.00003
IJI RES_LOW	-0.10498	-0.31395	0.04246	-0.03410	0.06421	-0.01151	0.01306	0.00078	0.00103
LAND_NumP RES_LOW	-0.04204	0.00561	-0.02993	0.10641	0.09619	0.12893	0.01109	-0.01651	0.00209
LAND_PSCoV RES_LOW	-0.25506	0.11678	-0.15841	-0.07126	-0.01165	-0.03423	0.01289	0.00159	-0.00342
LAND_PSSD RES_LOW	0.03920	-0.02483	0.05232	-0.07409	-0.05731	0.10163	0.01935	0.00378	0.00003
LAND_TE RES_LOW	0.03605	-0.01873	0.03496	0.00647	0.01225	-0.04488	0.00514	0.05276	-0.00459
LAND_ED RES_LOW	0.19631	0.04053	0.07014	-0.00203	-0.05591	0.02288	0.05808	0.00033	-0.00047
LAND_MSI RES_LOW	-0.15257	0.01686	0.24634	0.01862	-0.01516	0.00820	-0.01728	-0.00054	-0.00365
LAND_AWMSI RES_LOW	0.15278	-0.04153	-0.11809	-0.12212	0.12603	-0.01119	-0.00740	-0.00354	0.00512
LAND_MPFD RES_LOW	-0.01553	0.00925	-0.07915	0.09986	-0.09780	0.00082	0.02735	0.00907	0.03396
LAND_AWMPFD RES_LOW	0.04434	-0.01776	-0.10294	0.08204	-0.07351	0.00584	0.01486	-0.00276	-0.03616
LPI RES_LOW	0.03705	0.03033	-0.00837	0.16941	0.08869	-0.06262	-0.03654	-0.00017	0.00026
LSI RES_LOW	0.05645	-0.00830	-0.00099	-0.04257	-0.08536	0.00361	-0.14581	-0.00900	0.00321
LAND_CA RES_LOW	0.02596	-0.02849	0.05938	-0.01180	-0.04632	-0.11831	0.05131	-0.03474	0.00233
DLFD RES_LOW	0.01942	0.21965	0.10311	-0.06245	0.10580	-0.01074	0.00831	-0.00063	0.00034
LAND_MPAR RES_LOW	0.07144	-0.01835	-0.06749	-0.00493	0.00658	0.00089	0.00234	0.00013	0.00077

Figure 61. Eigen values and Loading Matrix for Level 1 Low Density Residential Index

PCA Index – RES_HIGH

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	7.3153	40.641	40.641
2	2.5444	14.135	54.776
3	1.9228	10.682	65.458
4	1.3266	7.370	72.828
5	0.8743	4.857	77.685
6	0.7948	4.416	82.101
7	0.7414	4.119	86.220
8	0.6986	3.881	90.101
9	0.6314	3.508	93.609
10	0.4927	2.737	96.346
11	0.2257	1.254	97.600
12	0.1858	1.032	98.632
13	0.0921	0.512	99.143
14	0.0648	0.360	99.503
15	0.0418	0.232	99.736
16	0.0314	0.174	99.910
17	0.0138	0.077	99.986
18	0.0025	0.014	100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS RES_HIGH	0.48295	-0.20631	0.48111	-0.36615	-0.27384	-0.37001	0.18645	0.02313	0.14553
MPI RES_HIGH	0.43226	-0.23595	0.23150	0.33893	0.33647	0.13043	-0.10880	0.61233	0.21095
MNN RES_HIGH	0.26121	0.10396	-0.46574	-0.32025	-0.43755	0.32477	0.35167	0.39894	-0.01311
IJI RES_HIGH	0.88178	-0.08089	-0.06261	-0.21353	-0.03980	0.06616	-0.09465	-0.04175	0.15846
LAND_NumP RES_HIGH	0.75880	-0.04315	-0.45173	0.28843	0.03680	-0.14625	-0.01876	-0.08144	0.10063
LAND_PSCoV RES_HIGH	0.75014	-0.00215	-0.30479	0.18447	0.02112	0.37138	-0.09224	-0.19503	-0.15581
LAND_PSSD RES_HIGH	0.70271	-0.24731	0.13068	0.26537	-0.12556	0.18293	0.13618	-0.07301	-0.37598
LAND_TE RES_HIGH	0.83479	-0.22479	-0.17656	0.32129	-0.04497	-0.25563	0.10062	-0.00542	0.03836
LAND_ED RES_HIGH	0.63093	-0.23013	0.54404	-0.10372	0.11943	0.22835	-0.08059	-0.05160	-0.01415
LAND_MSI RES_HIGH	0.31972	0.85189	0.24233	0.17379	-0.04200	-0.07021	-0.00427	0.09118	-0.18739
LAND_AWMSI RES_HIGH	0.26275	0.83214	0.28880	0.16769	-0.05814	-0.16519	0.00293	0.16091	-0.20018
LAND_MPFD RES_HIGH	0.79203	0.41499	-0.12678	-0.27489	0.02168	0.00557	-0.23944	-0.03144	0.04559
LAND_AWMPFD RES_HIGH	0.77807	0.42536	-0.10853	-0.29185	0.02109	-0.04102	-0.22688	-0.00131	0.08866
LPI RES_HIGH	0.54053	-0.27499	0.69619	-0.17377	-0.04504	0.17801	0.04180	-0.06585	-0.09914
LSI RES_HIGH	0.92334	-0.04305	-0.14951	-0.16611	0.04084	-0.01028	-0.08697	-0.04791	0.11834
LAND_CA RES_HIGH	0.80770	-0.29274	-0.03213	0.32782	-0.12730	-0.27519	0.18430	0.00419	-0.03240
DLFD RES_HIGH	0.34171	0.06877	-0.20036	-0.39741	0.65663	-0.10538	0.44398	0.00381	-0.20081
LAND_MPAR RES_HIGH	0.06591	0.56835	0.23982	0.29254	0.04620	0.26119	0.41095	-0.26015	0.46117

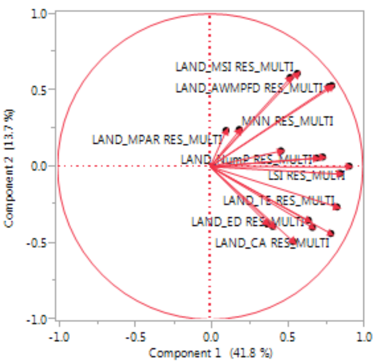
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS RES_HIGH	-0.20179	0.18001	0.10282	-0.05656	0.03146	0.04900	-0.00056	-0.00176	-0.00004
MPI RES_HIGH	-0.15798	0.03535	0.03483	0.00658	-0.00300	0.00492	-0.00179	-0.00036	0.00030
MNN RES_HIGH	0.14021	-0.03517	0.02411	-0.00904	0.00207	0.00042	-0.00244	0.00080	0.00024
IJI RES_HIGH	0.01346	0.16449	-0.27081	0.14426	0.05536	-0.01296	0.00446	-0.00097	0.00047
LAND_NumP RES_HIGH	0.23486	0.04814	0.12210	0.08400	-0.03975	0.11218	-0.02591	0.03133	-0.00101
LAND_PSCoV RES_HIGH	-0.09821	0.24775	0.14942	-0.05657	0.02893	-0.03988	0.04621	-0.00582	-0.00262
LAND_PSSD RES_HIGH	-0.33196	-0.12089	-0.10362	0.00956	0.00087	0.08297	-0.02324	-0.00349	-0.00005
LAND_TE RES_HIGH	0.15863	-0.06431	-0.00545	-0.02572	0.02383	-0.02302	0.00461	-0.08983	0.00377
LAND_ED RES_HIGH	0.36733	-0.06530	-0.03111	-0.11946	0.10169	0.03673	-0.00565	0.01743	-0.00009
LAND_MSI RES_HIGH	0.06516	0.11658	-0.00679	-0.00980	0.00643	-0.03732	-0.11605	-0.00383	-0.00955
LAND_AWMSI RES_HIGH	0.10382	0.03267	-0.06357	0.00109	-0.02961	0.03310	0.11117	0.00610	0.00861
LAND_MPFD RES_HIGH	-0.13645	-0.12760	0.09602	0.01536	0.03685	-0.01527	-0.02781	0.00344	0.03455
LAND_AWMPFD RES_HIGH	-0.12033	-0.18138	0.06615	0.02017	0.03077	0.00767	0.03069	-0.00859	-0.03281
LPI RES_HIGH	0.12696	-0.03299	0.12796	0.13946	-0.11225	-0.04339	0.00433	-0.01295	-0.00028
LSI RES_HIGH	-0.01429	0.00784	-0.13028	-0.14953	-0.17327	-0.01311	-0.01111	0.00919	-0.00018
LAND_CA RES_HIGH	-0.01609	-0.09549	0.00157	0.00243	0.03764	-0.10705	0.01455	0.06310	-0.00212
DLFD RES_HIGH	-0.00416	0.01986	0.00470	0.01240	0.00993	-0.00240	0.00202	-0.00033	-0.00002
LAND_MPAR RES_HIGH	-0.08551	-0.05064	-0.00526	-0.00115	0.00028	0.00419	0.00479	-0.00101	0.00127

Figure 62. Eigen values and Loading Matrix for Level 1 High Density Residential Index

PCA Index – RES_MULTI

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	7.5270	41.817	41.817
2	2.4675	13.708	55.525
3	1.6697	9.276	64.801
4	1.2394	6.886	71.687
5	0.9559	5.310	76.997
6	0.8666	4.814	81.812
7	0.7580	4.211	86.023
8	0.7221	4.012	90.034
9	0.5942	3.301	93.335
10	0.4127	2.293	95.628
11	0.2583	1.435	97.063
12	0.1541	0.856	97.919
13	0.1260	0.700	98.619
14	0.0977	0.543	99.162
15	0.0710	0.394	99.556
16	0.0640	0.356	99.911
17	0.0133	0.074	99.985
18	0.0026	0.015	100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS RES_MULTI	0.41601	-0.39001	0.49484	-0.46365	0.10358	0.11633	0.05693	-0.30743	0.12078
MPI RES_MULTI	0.37441	-0.37312	0.12694	0.24952	-0.24577	0.42414	0.29119	0.43414	0.35398
MNN RES_MULTI	0.19286	0.23768	-0.34830	-0.42815	0.46225	0.45234	-0.29047	0.23867	0.04946
IJI RES_MULTI	0.85503	-0.04917	-0.18615	-0.17082	0.10195	-0.13986	-0.04473	-0.02604	0.14181
LAND_NumP RES_MULTI	0.70206	0.05527	-0.53322	0.21546	-0.12432	-0.13028	-0.00956	-0.14557	0.16270
LAND_PSCoV RES_MULTI	0.73904	0.05893	-0.34741	0.33693	0.09011	-0.05692	-0.17652	0.17710	-0.17693
LAND_PSSD RES_MULTI	0.67322	-0.39988	0.04662	0.17104	0.07989	0.28546	0.00585	0.01405	-0.45012
LAND_TE RES_MULTI	0.83226	-0.26403	-0.30388	0.09225	-0.09437	0.06129	0.07089	-0.24723	0.02677
LAND_ED RES_MULTI	0.64371	-0.35300	0.25453	0.11141	0.01529	-0.38880	-0.25968	0.25710	0.08997
LAND_MSI RES_MULTI	0.57122	0.60760	0.34713	0.25458	0.00233	0.08588	-0.00686	0.02239	-0.12836
LAND_AWMSI RES_MULTI	0.52416	0.57998	0.41894	0.02059	-0.28145	0.15383	-0.04107	-0.04433	-0.09608
LAND_MPFD RES_MULTI	0.80298	0.52519	0.09840	-0.10660	-0.08023	-0.01322	-0.02237	0.01418	0.05224
LAND_AWMPFD RES_MULTI	0.78009	0.52173	0.11691	-0.17504	-0.12214	0.00188	-0.01235	-0.02261	0.09616
LPI RES_MULTI	0.54710	-0.49218	0.51766	-0.03160	0.13251	-0.16258	-0.23330	0.15021	-0.03918
LSI RES_MULTI	0.91005	-0.00120	-0.14724	-0.16882	0.01886	-0.10042	-0.00527	-0.03627	0.14295
LAND_CA RES_MULTI	0.79537	-0.44092	-0.06945	0.02712	-0.02646	0.20392	0.10336	-0.25701	-0.11197
DLFD RES_MULTI	0.46975	0.10311	-0.10021	-0.37167	0.15484	-0.28790	0.60164	0.28359	-0.22899
LAND_MPAR RES_MULTI	0.10519	0.23584	0.25415	0.51377	0.69166	-0.01317	0.22432	-0.16721	0.19465

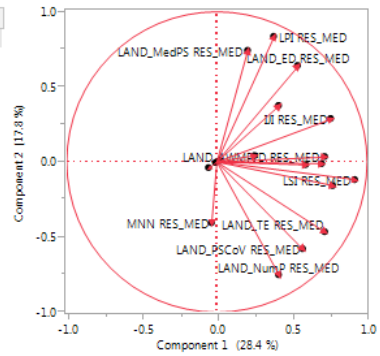
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS RES_MULTI	-0.04130	0.24415	-0.01068	-0.04880	0.11913	0.01562	0.00888	0.00747	-0.00013
MPI RES_MULTI	-0.04807	0.02859	-0.01157	0.00769	0.00533	-0.00078	0.00128	0.00105	-0.00011
MNN RES_MULTI	0.19853	-0.00681	0.00588	-0.00340	-0.00305	0.01008	-0.00189	-0.00060	-0.00026
IJI RES_MULTI	-0.21030	-0.10160	-0.28724	0.11277	0.00706	0.00415	-0.00882	-0.00379	-0.00049
LAND_NumP RES_MULTI	0.18934	0.13371	0.06702	0.11986	0.05528	0.10582	0.07616	-0.02215	0.00008
LAND_PSCoV RES_MULTI	-0.17055	0.23937	0.01692	-0.02373	0.08348	-0.12189	-0.04326	-0.00039	0.00333
LAND_PSSD RES_MULTI	-0.15885	-0.14139	0.02756	0.00786	0.06884	0.09845	0.06365	0.01440	-0.00060
LAND_TE RES_MULTI	0.20327	-0.03994	-0.01385	-0.03042	-0.05163	-0.04055	-0.04594	0.08239	-0.00239
LAND_ED RES_MULTI	0.20858	-0.11566	-0.02435	-0.14546	0.08744	0.05239	-0.04150	-0.00731	-0.00013
LAND_MSI RES_MULTI	0.07415	0.19100	-0.14399	-0.09040	-0.12972	0.08183	0.01874	0.00172	0.00715
LAND_AWMSI RES_MULTI	0.22685	-0.10447	-0.04902	0.06989	0.10335	-0.11373	0.06160	-0.00867	-0.00593
LAND_MPFD RES_MULTI	-0.16103	-0.01049	0.10737	0.00415	-0.01871	0.05168	-0.07998	0.00140	-0.03621
LAND_AWMPFD RES_MULTI	-0.12798	-0.09853	0.12671	0.03203	0.02026	0.02882	-0.06624	0.00950	0.03500
LPI RES_MULTI	0.04291	0.06454	0.08720	0.18851	-0.12717	-0.03242	0.02237	0.00902	0.00043
LSI RES_MULTI	-0.14262	-0.07198	0.07211	-0.15140	-0.08443	-0.06778	0.15467	-0.01100	-0.00067
LAND_CA RES_MULTI	0.10774	-0.04982	0.00640	-0.02522	-0.07492	-0.03552	-0.09445	-0.07281	0.00206
DLFD RES_MULTI	0.13012	0.03728	0.00623	0.01784	0.00965	-0.00741	-0.00150	0.00031	0.00033
LAND_MPAR RES_MULTI	0.00658	-0.08301	0.03675	0.01642	0.03272	-0.02024	0.00056	0.00005	-0.00096

Figure 63. Eigen values and Loading Matrix for Level 1 Multi Family Residential Index

Level 1 PCA Index – RES_MED

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	5.1131	28.406		28.406
2	3.1988	17.771		46.177
3	2.8315	15.731		61.908
4	1.7545	9.747		71.655
5	1.1577	6.432		78.087
6	0.9340	5.189		83.276
7	0.8708	4.838		88.113
8	0.5780	3.211		91.325
9	0.4515	2.508		93.833
10	0.2664	1.480		95.313
11	0.2440	1.356		96.669
12	0.2188	1.216		97.885
13	0.1638	0.910		98.795
14	0.0791	0.439		99.234
15	0.0605	0.336		99.570
16	0.0383	0.213		99.783
17	0.0227	0.126		99.910
18	0.0163	0.090		100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS RES_MED	0.20597	0.74138	-0.02791	0.14348	0.50601	0.12400	-0.15336	0.04815	-0.16915
MPI RES_MED	0.41075	0.37100	-0.05124	0.41891	-0.02558	0.41288	-0.02229	-0.50891	0.29003
MNN RES_MED	-0.03477	-0.40879	0.00543	-0.41271	0.18620	0.65068	-0.25009	0.29634	0.19407
IJI RES_MED	0.75853	0.28388	-0.04113	-0.31385	-0.00746	-0.07996	-0.04630	0.17099	0.23239
LAND_NumP RES_MED	0.41517	-0.75483	-0.05941	0.09692	0.32217	-0.20204	-0.01267	-0.04189	0.11446
LAND_PSCoV RES_MED	0.57081	-0.57916	-0.00648	0.00710	-0.45946	0.03685	0.01824	0.02389	0.03832
LAND_PSSD RES_MED	0.59089	-0.01940	-0.04538	0.34401	-0.56083	0.31733	0.06597	0.14079	-0.22797
LAND_TE RES_MED	0.72197	-0.46819	-0.09650	0.35948	0.29300	-0.08566	0.02470	0.04949	-0.00721
LAND_ED RES_MED	0.54139	0.63719	-0.01274	-0.08524	-0.18381	-0.27305	0.11224	0.14160	0.25897
LAND_MSI RES_MED	-0.00499	-0.00792	0.97672	0.12805	0.01918	0.00490	0.02057	0.03384	0.02400
LAND_AWMSI RES_MED	0.01950	0.00716	0.93883	0.15951	0.03466	0.01497	0.02671	0.05108	0.03429
LAND_MPFD RES_MED	0.70143	-0.01202	0.26743	-0.54532	-0.05948	-0.00440	-0.18564	-0.21685	-0.19493
LAND_AWMPFD RES_MED	0.71747	0.03168	0.26341	-0.53773	0.03878	-0.02464	-0.17390	-0.21220	-0.16066
LPI RES_MED	0.38026	0.83049	-0.02699	0.09125	-0.03924	0.04468	-0.06379	0.20748	-0.04850
LSI RES_MED	0.92184	-0.11739	-0.06639	0.02556	0.05747	-0.16190	0.02852	0.03970	0.10330
LAND_CA RES_MED	0.77250	-0.15682	-0.10586	0.47782	0.24399	0.08522	-0.01058	0.10853	-0.17015
DLFD RES_MED	0.25013	0.04256	0.00349	-0.36908	0.21462	0.23928	0.83097	-0.02139	-0.05920
LAND_MPAR RES_MED	-0.05272	-0.04331	0.90480	0.15309	0.01148	-0.00425	0.05155	0.04864	0.04919

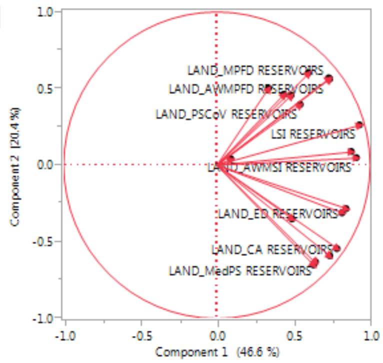
	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS RES_MED	0.10853	0.01880	-0.01572	0.17984	0.10007	0.07351	0.02154	0.00888	0.00143
MPI RES_MED	0.01360	0.00227	0.01206	-0.01923	-0.00768	-0.00502	-0.00267	-0.00044	-0.00076
MNN RES_MED	-0.09535	0.07580	-0.01940	-0.00233	0.00021	0.00234	-0.00399	-0.00128	-0.00062
IJI RES_MED	0.22842	-0.30903	-0.00088	-0.05391	-0.02791	0.02531	0.00191	0.00258	-0.00206
LAND_NumP RES_MED	0.10254	0.10627	0.16028	-0.13625	0.12919	0.03207	0.00238	-0.03628	0.01477
LAND_PSCoV RES_MED	0.20099	0.13754	0.09056	0.23170	-0.04076	0.01199	0.02573	0.00945	0.00815
LAND_PSSD RES_MED	-0.05203	-0.08720	-0.02997	-0.10583	0.11931	0.04406	0.00824	0.00534	0.00310
LAND_TE RES_MED	-0.08085	0.00625	-0.01329	-0.03928	-0.05537	0.02293	0.01508	0.10623	-0.03628
LAND_ED RES_MED	-0.20418	0.17311	-0.02651	0.00106	-0.00857	0.11252	0.00915	-0.01438	0.00388
LAND_MSI RES_MED	-0.02291	-0.03686	0.07338	0.03145	0.00119	0.02540	-0.11766	0.03530	0.05484
LAND_AWMSI RES_MED	-0.13901	-0.12056	0.19881	0.05771	0.00281	-0.02339	0.08350	-0.02999	-0.04059
LAND_MPFD RES_MED	0.00081	0.05502	0.01564	-0.01870	-0.00697	0.03549	-0.07599	-0.01980	-0.06937
LAND_AWMPFD RES_MED	-0.07709	0.01286	-0.02150	-0.05370	-0.02110	-0.01997	0.08550	0.02153	0.06651
LPI RES_MED	0.10438	0.18481	0.17618	-0.10874	-0.02233	-0.12150	-0.01133	0.01761	-0.00435
LSI RES_MED	-0.13028	-0.04082	-0.14838	0.13026	0.09804	-0.14215	-0.03371	-0.01469	-0.00060
LAND_CA RES_MED	-0.02671	-0.01620	-0.03740	-0.00962	-0.14786	0.01528	-0.01813	-0.07836	0.02133
DLFD RES_MED	0.02509	0.01765	0.02500	0.00535	-0.00026	-0.00346	-0.00110	0.00049	-0.00036
LAND_MPAR RES_MED	0.19065	0.13834	-0.28581	-0.08225	0.00276	-0.00853	0.03612	-0.00766	-0.01634

Figure 64. Eigen values and Loading Matrix for Level 1 Medium Density Residential Index

PCA Index - RESERVOIRS

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	8.3930	46.628	46.628
2	3.6809	20.450	67.077
3	1.4650	8.139	75.216
4	1.0359	5.755	80.971
5	0.8013	4.451	85.423
6	0.6978	3.877	89.300
7	0.5701	3.167	92.467
8	0.4244	2.358	94.825
9	0.3174	1.763	96.588
10	0.2214	1.230	97.818
11	0.1651	0.917	98.735
12	0.0907	0.504	99.239
13	0.0690	0.384	99.623
14	0.0323	0.180	99.802
15	0.0195	0.108	99.910
16	0.0096	0.053	99.964
17	0.0061	0.034	99.997
18	0.0005	0.003	100.000



Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
LAND_MedPS RESERVOIRS	0.64407	-0.63648	-0.15433	-0.13056	-0.02198	0.29456	0.06000	0.12250	0.05984
MPI RESERVOIRS	0.63555	-0.65582	0.05458	-0.09563	0.17691	0.04004	0.09145	0.21419	0.08331
MNN RESERVOIRS	0.33904	0.50246	0.08212	-0.15370	0.55329	0.41851	-0.27976	-0.17080	0.09612
IJI RESERVOIRS	0.59880	0.59733	0.07883	-0.09787	0.03704	0.12995	-0.01174	0.31013	-0.34269
LAND_NumP RESERVOIRS	0.43817	0.46116	0.56718	0.01124	-0.32822	0.18728	0.18403	0.06659	0.20009
LAND_PSCoV RESERVOIRS	0.54675	0.39573	0.60671	0.13881	-0.09810	0.01214	-0.06825	-0.14057	-0.15816
LAND_PSSD RESERVOIRS	0.49275	-0.35046	0.41049	0.09526	0.43386	-0.48425	-0.03069	0.05564	-0.03482
LAND_TE RESERVOIRS	0.84986	-0.28414	0.35190	0.01794	-0.11588	0.01321	0.02759	-0.12219	0.07393
LAND_ED RESERVOIRS	0.81719	-0.31147	-0.21554	0.03368	-0.17235	0.03385	-0.05269	-0.20713	-0.23171
LAND_MSI RESERVOIRS	0.87922	0.08579	-0.34476	0.10319	-0.11810	-0.05042	-0.11816	-0.09188	0.10097
LAND_AWMSI RESERVOIRS	0.91560	0.04626	-0.11284	0.06320	-0.07941	-0.19146	-0.14692	-0.19647	0.06256
LAND_MPFD RESERVOIRS	0.73337	0.56516	-0.25843	-0.00014	-0.01204	-0.13872	-0.05407	0.16657	0.09969
LAND_AWMPFD RESERVOIRS	0.73859	0.56285	-0.25010	-0.01624	-0.01354	-0.14606	-0.05526	0.15994	0.09825
LPI RESERVOIRS	0.73794	-0.59118	-0.14731	-0.00926	-0.03139	0.16852	-0.00089	-0.02037	-0.12988
LSI RESERVOIRS	0.93340	0.26044	-0.08950	-0.03758	-0.05730	-0.05522	-0.03558	0.03739	0.00400
LAND_CA RESERVOIRS	0.78869	-0.54459	0.21640	-0.02809	0.06649	0.04264	0.05080	0.05940	0.05224
DLFD RESERVOIRS	0.48569	0.45350	-0.21062	-0.15456	0.24878	-0.03609	0.61668	-0.20884	-0.03712
LAND_MPAR RESERVOIRS	0.09216	0.03708	-0.12183	0.95159	0.14863	0.17897	0.10510	0.06130	0.00886

	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
LAND_MedPS RESERVOIRS	-0.09502	-0.00309	-0.09119	0.07516	-0.05774	0.03362	0.04556	0.01511	-0.00031
MPI RESERVOIRS	-0.08395	0.01512	0.21986	-0.06237	-0.00477	-0.01509	-0.00155	0.00424	0.00022
MNN RESERVOIRS	0.06242	0.01274	0.01142	-0.00598	0.00375	0.00430	-0.00093	-0.00082	-0.00003
IJI RESERVOIRS	0.02982	-0.18928	0.00122	0.00439	0.00920	-0.00039	-0.00000	-0.00060	-0.00015
LAND_NumP RESERVOIRS	0.21446	0.01661	0.04543	0.06104	0.00134	-0.01048	0.01177	-0.00857	-0.00052
LAND_PSCoV RESERVOIRS	-0.27903	0.13038	0.03133	0.04113	-0.01817	-0.00028	0.00020	0.00579	0.00050
LAND_PSSD RESERVOIRS	0.14692	0.01403	-0.03090	0.07266	-0.02453	0.00256	0.01019	0.00943	0.00021
LAND_TE RESERVOIRS	0.01357	-0.12477	-0.06710	-0.11859	0.03288	0.02510	-0.01835	0.04145	0.00131
LAND_ED RESERVOIRS	0.18968	0.09636	0.08243	-0.02386	-0.01943	0.06543	0.00433	-0.00856	-0.00029
LAND_MSI RESERVOIRS	-0.02213	-0.13190	0.05019	0.12339	-0.05329	-0.01219	-0.04806	0.00882	0.00233
LAND_AWMSI RESERVOIRS	-0.06366	-0.15570	0.03178	-0.00860	0.04678	-0.02659	0.05521	-0.02044	-0.00282
LAND_MPFD RESERVOIRS	-0.02815	0.11160	-0.01254	0.00310	0.03489	0.02928	-0.00978	0.01346	-0.01473
LAND_AWMPFD RESERVOIRS	-0.02453	0.10454	-0.01576	-0.00843	0.04629	0.02772	0.00923	0.00028	0.01504
LPI RESERVOIRS	0.06508	0.12257	-0.04447	0.05662	0.08286	-0.07725	-0.00878	0.01205	0.00032
LSI RESERVOIRS	0.04257	0.07720	-0.07386	-0.12201	-0.10253	-0.06132	0.00006	-0.01060	-0.00020
LAND_CA RESERVOIRS	-0.07385	-0.01704	-0.08493	-0.00230	0.02645	0.02950	-0.03556	-0.05379	-0.00088
DLFD RESERVOIRS	-0.03825	-0.01132	-0.00173	0.01078	0.00167	-0.00142	-0.00051	0.00067	0.00001
LAND_MPAR RESERVOIRS	0.00829	-0.00965	-0.00608	-0.02292	-0.00041	0.00190	0.00543	-0.00123	0.00002

Figure 65. Eigen values and Loading Matrix for Level 1 Reservoirs Index